Universidad Carlos III de Madrid
Escuela Politécnica Superior

Departamento de Informática
Doctorado en Ciencia y Tecnología Informática

Tesis Doctoral

A General Cognitive Framework for Contex-Aware Systems: Extensions and Applications for High-Level Information Fusion Approaches

Miguel Ángel Serrano Mateos

Dirigida por

Miguel Ángel Patricio

Jesús García

Madrid, 2015

E-mail: miguel.serrano@uc3m.es

Telephone: +34 648088629

Address:

Grupo de Inteligencia Artificial Aplicada
Departamento de Informática
Universidad Carlos III de Madrid
Av. de la Universidad Carlos III, 22
Colmenarejo 28270 — Spain

A General Cognitive Framework for Contex-Aware Systems: Extensions and Applications for High-Level Information Fusion Approaches

Autor: Miguel A. Serrano
Directores: Miguel A. Patricio
Jesús García

Firma del Tribunal Calificador:

	Nombre y Apellidos	Firma
Presidente: D.	....	....
Vocal: D.	....	....
Vocal: D.	....	....
Vocal: D.	....	....
Secretario: D.	....	....

Calificación: ....

Colmenarejo, ... de .... de 2011.

A mis padres, a mis hermanos, a mi familia.

Contents

List of Figures

List of Tables

Abstract

Context-aware systems aims at the development of computational systems that process data acquired from different datasources and adapt their behaviour in order to provide the 'right' information, at the 'right' time, in the 'right' place, in the 'right' way to the 'right' person (Fischer, 2012). Traditionally computational research has tried to answer these needs by means of low-level algorithms. In the last years the combination of numeric and symbolic approaches has offered the opportunity to create systems to deal with these issues. However, although the performance of algorithms and the quality of the data directly provided by computers and devices has quickly improved, symbolic models used to represent the resulting knowledge have not yet been adapted to smart environments. This lack of representation does not allow to take advantage of the semantic quality of the information provided by new sensors.

This dissertation proposes a set of extensions and applications focused on a cognitive framework for the implementation of context-aware systems based on a general model inspired by the Information Fusion paradigm. This model is stepped in several abstraction levels from low-level raw data to high level scene interpretation whose structure is determined by a set of ontologies. Each ontology level provides a skeleton that includes general concepts and relations to describe entities and their connections. This structure has been designed to promote extensibility and modularity, and might be refined to apply this model in specific domains. This framework combines a priori context knowledge represented with ontologies with real data coming from sensors to support logic-based high-level interpretation of the current situation and to automatically generate feedback recommendations to adjust data acquisition procedures.

This work advocates for the introduction of general purpose cognitive layers in order to obtain a closer representation to the human cognition, generate additional knowledge and improve the high-level interpretation. Extensibility and adaptability of the basic ontology levels is demonstrated with the introduction of these traverse semantic layers which are able to be present and represent information at several granularity levels of knowledge using a common formalism.

Context-based system must be able to reason about uncertainty. However the reasoning associated to ontologies has been limited to classical description logic mechanisms. This research also tackle the problem of reasoning under uncertainty circumstances through a logic-based paradigm for abductive reasoning: the Belief-Argumentation System.

The main contribution of this dissertation is the adaptation of the general architecture and the theoretical proposals to several context-aware application areas such as Ambient Intelligence, Social Signal Processing and surveillance systems. The implementation of prototypes and examples for these areas are explained along this dissertation to progressively illustrate the improvements and extensions in the framework. To initially depict the general

model, its components and the basic reasoning mechanisms a video-based Ambient Intelligence application is presented. The advantages and features of the framework extensions through traverse cognitive layers are demonstrated in a Social Signal Processing case for the elaboration of automatic market researches. Finally, the functioning of the system under uncertainty circumstances is illustrated with several examples to support decision makers in the detection of potential threats in common harbor scenarios.

Agradecimientos

Sin

Miguel Angel Serrano
Madrid, enero 2016.

1

Introduction

1.1 Motivaton

During the last decades the world has suffered a frantic technological transformation due to great advances in hardware and the generalization of the economies of scale. The emergence of a huge amount of new devices, their availability in the market at low price and their capacity to generate relevant data have fueled future expectations in context-aware application areas such as Ambient Intelligence (AmI), Social Signal Processing (SSP), Ambient Assisted Living (AAL) and others which were frequently limited by the lack of contextual information.

All along these years, most disciplines involved in computer sciences have performed a huge effort to improve the efficiency and the accuracy of these systems using low level algorithms. Typically, the resulting data strongly based on numerical methods did not provide support to obtain, understand and manage semantically enriched or high level data. On the other hand, approaches focused on knowledge have traditionally tried to extract semantic descriptions using a higher abstraction perspective, however, none of these approaches could only by itself cover the entire problem. The growing number of context data sources such as, devices, networks, embedded sensors, users and so forth, require novel abilities to represent and retrieve information of interest.

Increasingly low level and knowledge approaches are being used in a synergistic way taking into account their abstraction levels. However, these mixed approaches are not normally prepared to accept different kinds of measures form different kinds of devices, to fuse all the data generated and to make inference reasoning over these. In addition, normally with these solutions it is not possible to deal with specific problems at different abstraction levels specially those refered to higher levels.

Data and Information Fusion (DIF) architectures have always advocated for systems organized in different abstraction levels. In these levels the detection and identification of real world entities and characterization of activities and threats require assessing the states of situational items and their relationships within a specific context. From the perspective of fusion processes, context can be informally defined as the set of background circumstances that are not of prime interest to the system, but have potential relevance towards optimal estimation (García et al., 2012). When a context is activated (i.e., some circumstances hold), more information is available to obtain and improve estimations on entities. This contextual information, expressed in the form of complementary knowledge or constraints, encompasses information about objects, processes, events, and relationships between them, as well as particular goals, plans, capabilities, and policies of the decision makers. Such diversity makes formal context representation a significant challenge.

In order to handle such a diverse dataset is necessary to provide well-defined models to represent the context and perceptual semantics in complex scenarios and also provide and adequate formalism compatible with any reasoning mechanisms. Current symbolic data representations allow to develop cognitive models able to represent accurately the scenes' complexity and perform inferences. According to Vernon's definition "Cognitivism asserts that cognition involves computations defined over symbolic representations, in a process whereby information about the world is abstracted by perception, represented using some appropriate symbol set, reasoned about, and then used to plan and act in the world." (Vernon, 2008) These models can analyze systematically the knowledge of the scene to discover and describe data related with activities developed by a subject fusing its representation with high-level context knowledge. In this context, the use of ontology models offer several advantages at a low cost. Formal models establish a common symbolic vocabulary to describe and communicate scene data while providing support for logic-based reasoning. Symbolic language is closer to human language, and therefore it is easy to interact and interpret system inputs and outputs. Logic-based reasoning, in turn, can be applied to check the consistency of the models and to infer additional knowledge from explicit information.

Despite the advantages of ontology models, the state of the art in this technology presents serious drawbacks both in the availability of standards, languages, libraries and developing tools as in the reasoning capabilities, mainly as a consequence of the open world assumption –any statement that is not known to be true is undefined on unknown. Along the last years these disadvantages are being mitigated by the World Wide Web Consortium (W3C) which periodically publish new drafts and recommendations and by educational institutions, companies and open source projects which provide support for modeling tools, reasoners, libraries and so forth. Furthermore, the possibility of endow new features to the reasoning procedures bypassing the limitations of descriptive logics has encouraged researchers to start new challenges regarding reasoners and reasoning techniques. Some works have been carrying out at different levels, from the redefinition of description logics based on artificial intelligence techniques such as fuzzy logic, to the imbibe of Bayesian probability theory to the classical first-order logic.

Beacuse of theses drawbacks this work aims to outline a dual purpose: (i) explore the implementation of applications on the basis of an existing general cognitive framework provided with reasoning capabilities by means of Artificial Intelligence and DIF techniques; (ii) update, create and promote knowledge representation layers in in order to extend the representation and reasoning capabilities of context-aware systems.

1.2 Thesis proposal

The use of situation-aware systems aims to automatically detect the behavior of the detected objects in a scene. Providing additional information to this system, namely, context knowledge, make this task more feasible also introducing certain complexities. Context knowledge comprises any external knowledge used to complete the information about the computed scene, among other categories context includes: (i) environment of the scene, such as structures, static objects or behavior features; (ii) devices' parameters; (iii) historical records; (iv) data coming from human-computer interaction. The classical techniques –based on data coming from observations and a priori knowledge models– have been insufficient to recognize situations in complex and unpredictable scenarios.

On the other hand, sensor fusion architectures are commonly known to provide data organized in multi-level structures. Despite this valuable way of organizing data, most researches in the sensor fusion literature have only taken into account the measures of the local context to achieve a more accurate understanding of the scene. This circumpstance makes these methods barely extensible to other domains since oftentimes perform application dependent and heuristics calculations.

In contrast to these approaches, multi-level cognitive approaches propose symbolic models at different abstraction levels which define the domain in a logic language able to represent the semantics of objects and relationships in the environment. As cognitive practical approaches are scarce or nonexistent this thesis proposes the construction of a set of applications based on a general purpose cognitive framework inspired in a symbolic multi-level architecture. The architecture of this system is based in the ontological model based on DIF presented in (Gómez-Romero et al., 2011a). The objective of this dissertation is predominantly practical. The approaches presented in this thesis will offer detailed explanation of the implementation issues and practical demonstrations executed over developed prototypes. Due to the increasing amount of sensor technologies it has been decided to design the solutions under the ontology standards of generality, modularity and extensibility. Compliance with these requirements, implementations are demonstrated using different datasources in a variety of application fields. The reasoning capabilities the framework will be progressively showed with new representation and reasoning capabilities. These capabilities will be demonstrated in different example scenarios dealing with complex situations.

1.2.1 Classification, rule-based and spatial reasoning with ontologies

This document will cover the current representation and reasoning needs in context-aware approaches through the ontological representation of general concepts. This representation is not just a conceptual map to store data coming from sensorized environmet systems, but also is a powerful tool to carry out reasoning processes. The standard reasoning procedure in ontologies can be apply to infer additional knowledge from explicit facts detected in the scene. Through the use of an inference engine, the subsumption mechanism can carry out classification tasks such as, determine the concept hierarchy or check the membership relation between instances and concepts.

Rules is the most extended deductive reasoning mechanism supported by ontologies. Standardized languages such as SWRL for rule-based reasoning with OWL ontologies, as well as, rule languages for specific reasoners such as the new RACER Query Language (nRQL) are tools able to maintain the consistency of the knowledge base and carry out individual classification tasks beyond the subsumption mechanism.

A huge quantity of knowledge that characterize situations lies in the spatial configuration of scene objects. Symbolic representation of spatial entities allow the use of reasoning formalisms based on qualitative spatial relationships. The result of these processes can be aggregated to an ontological model since reasoning and their outcomes is performed in linguistic terms.

1.2.2 Cognitive layers in knowledge representation

The formulation of models based on abstraction levels has led to the implementation of noncohesive systems which are not able to fluently communicate among themselves. For this reason, it is necessary to provide new common and transverse knowledge layers among these levels including new semantic relationships. The goal of this strategy is the close interaction among semantically similar layers to the automatic generation of new knowledge. With the advent of new sensors which will allow a more accurate detection of scene objects, we advocate for the addition of representation layers based on mereology and meronymy. The idea of employing a part-based layer to support the statements of the scene object abstraction level in a cognitive framwork has been previously suggested by Pinz et al. (Pinz et al., 2008). This proposal goes further and seeks to provide a symbolic layer based on the formal definition, development patterns and implementation of spatial and part-whole relationships.

Symbolic data representations allow to develop cognitive models able to represent more accurately the complexity of the scene. These models can analyze systematically the knowledge of the scene to discover and describe data related with activities developed by a subject fusing its representation with high-level context knowledge. A key part of such analysis is currently supported by the approaches emerged from a cognitive view of the traditional computer vision techniques. The ties between meronymy and the current qualitative approaches (Randell et al., 1992), (Allen, 1983) in cognitive vision –mainly focused on a qualitative description of spatio-temporal aspects (Renz, 2002)– must be regarded as crucial to narrow the gap of knowledge in context-based approaches.

1.2.3 Reasoning under uncertainty

Reasoning procedures presented in 1.2.1 are able to carry out classification or consistency checking tasks. However, there are no reasoning mechanisms associated to ontologies able to abduct new knowledge from the existing one and reason under uncertainty conditions. Abductive reasoning automatically infer suitable hypotheses that explain a set of input facts, in some cases, with degrees of uncertainty.

For the above reasons an extension of the reasoning capabilities of ontological approaches will be proposed with the goal of assessing situations in complex environments. These environments usually managed by people overwhelmed by the huge quantity of data coming from different sources and the duty of make decisions under pressure, require an approach which automatically recognizes anomalous situations and handles the derived threat with a level of certainty which support the decision making.

If we consider context and observed facts the inputs of the process and the situations explaining these facts the resulting hypotheses, the situation assessment can be understood as an abductive process. A Belief Argumentation Systems (BAS) (Rogova et al., 2006) –a logic-based paradigm used in combination with the Dempster-Shafer (DS) theory– can be used for this purpose. DS is an appropriate theory for uncertainty representation in this specific environment because generally in real-world complex situations the a priori probabilities of anomalous and threatening behaviors are not available. In addition this theory provides different mechanisms to fuse the hypotheses through combination rules.

1.3 Structure of the document

The rest of the document is organized as follows:

- Chapter 2: General architecture. Presentation of a multilevel cognitive framework inspired in high level Information Fusion models and guidelines.
- Chapter 3: Video based AmI. A starting approach for classification, rule-based and spatial reasoning applied to a CV smart home prototype.
- Chapter 4: Model extensions for video based AmI. Extension of previous proposals with additional cognitive layers presented using a market research case study.
- Chapter 5: Reasoning extensions. A proposal for reasoning under uncertainty with ontologies applied to a harbor surveillance prototype.
- Chapter 6: Conclusions and future works. Include a review of the main contributions of the document and a summary of the future trends of work.
- Apendix A: Context. Formal definitions of context for context-aware systems and the evolution of the context representation formalisms.
- Apendix B: Qualitative Spatial Representations (QRS). Review of the QRS formalisms applied to ontologies.
- Apendix C: Uncertainty and ontologies. General review of the approaches mixing any kind of uncertainty with ontologies. Special attention is paid to Bayesian Networks, Fuzzy Logic and Dempster-Shafer. An introduction to theoretical foundations of Belief Argumentation Systems (BAS) and Probabilistic Argumentation Systems (PAS) is also given due to its relevance to understand 5.

1.4 Note on thesis evaluation

The assessment of research works have to be founded on objective evaluation criteria. Most of the current frameworks and metrics try to measure the efficiency, accuracy, etc. of quantitative approaches while the description of the qualitative advantages is just mentioned. Since the approaches and models presented in this proposal are not just focused in quantitative aspects, as a general rule a qualitative evaluation will be performed for each proposal based on the new features and advantages added to the current thesis.

Assessment of high-level Information Fusion proposals is usually a task full of obstacles. Determine the quality of the results can be difficult since they can be based on non objective foundations. e.g. the application of subjective probabilities to decision making approaches. There is a lack of publicly available datasets mainly due to safety issues –the required information comes from sensors and infrastructures adhered to privacy protocols. The absence of common evaluation frameworks for high-level Information Fusion systems is being mitigated by the Evaluation of Techniques for Uncertainty Representation Working Group (ETURWG) [1], however, current proposals are not enough to make a reliable comparison of current systems.

The problem of the absence of common evaluation frameworks, datasets and implementations have been addressed differently depending on the situation. If available, proper evaluation methods, tools and public datasets have been used, otherwise prototypes and synthetic datasets have been built.

[1] Evaluation of Techniques for Uncertainty Representation Working Group (ETURWG). http://eturwg.c4i.gmu.edu/ Last accessed December 2015

2

General achitecture and core model

2.1 General architecture

This chapter present the general architecture shared by all the prototypes presented in this disertation. The overall architecture depicted in Fig. 2.1 is a general purpose knowledge approach for symbolic data representation coming from any source. The goal of the architecture is to carry out a comprehensive knowledge analysis extracting all the semantics at different abstraction levels in order to achieve a complete scene interpretation. Checking data consistency, creation of new knowledge and refinement tasks through inference processes are the tools employed to obtain a higher understanding of the scene. The architecture is divided in three layers; data adquisition, core model and reasoning procedures.

The data adquisition layer is resposible of receiving perceptual and contextual data coming directly from a sensor, multisensor and heterogeneous datasources; i.e., this layer collects data from raw resources and low level algorithms mainly circumscribed in the Information Fusion area.

The core model of the architecture presented is based on the Joint Directors of Laboratories (JDL) fusion model. This symbolic model is stepped in several levels from low-level track data to high level situations whose structure is determined by a set of ontologies (see 2.2).

The reasoning layer is divided in three features; an ontology based module, a cognitive layers reasoning module and a Belief Argumentation System (BAS) module.

The concept of ontology is equivalent to the representation of Description Logic (DL). Ontologies presented an optimum trade-off between representation and reasoning capabilities, adding support for consistency checking and subsumption reasoning mechanisms –given two concepts which one is more general– to its inherent features of expressiveness and interoperability. Usually reasoners also add support for abductive and deductive rule-based inference. Abductive rules defined in a sub-ontology create new instances in the same level or to an upper level. Deductive rules, are used to maintain the logical consistency of the scene.

Cognitive layers reasoning represent a key aspect for scene interpretation. These layers provide common and transverse knowledge among multi-level structures. Qualitative spatial relations and mereology provide a close interaction among semantically similar layers to the automatic generation of new knowledge. These layers allow to organize space in a more comprehensible way and also can serve as a basis to develop other semantic layers into the core model. A huge quantity of implicit relations can emerge from the combination of the knowledge associated to these layers.

BAS is a logic-based paradigm for abductive reasoning under uncertainty circumstances which encompasses the creation of hypotheses to explain the state of the world, the computation of the credibility of these hypotheses, and the selection of the most credible hypotheses C.4.

These reasoning processes allow the abduction of local and inter-level interpretations and the generation of low-level recomendations to update de configuration of the system behaviour in order to improve the quality of the data. The robustness of the data is guaranteed due to the reasoners consistency checking mechanism.

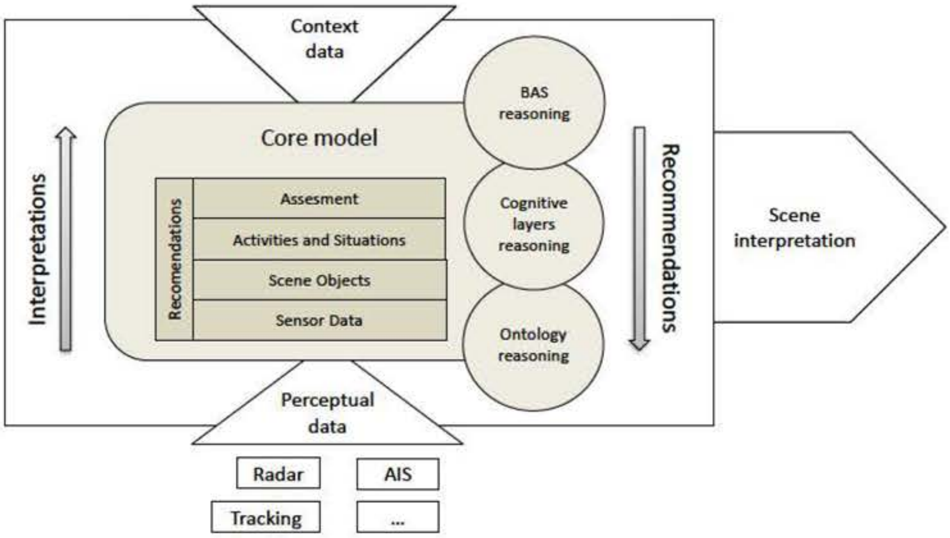


Figure 2.1: General architecture and core model

2.2 Core model: A JDL-based knowledge representation

Data and Information Fusion (DIF) research area studies theories and methods to effectively "combine data from multiple sensors and related information to achieve more specific inferences that could be achieved by using a single, independent sensor" (Hall and Llinas, 2009). The JDL model classifies fusion processes into five operational levels corresponding to different stages of the transformation from input signals to decision-ready knowledge (Steinberg and Bowman, 2004),(Llinas et al., 2004): signal feature assessment (L0), entity assessment (L1), situation assessment (L2), impact assessment (L3), and process assessment (L4).

During the last years, the first levels (L0, L1), aimed at the development of low level algoritms, has received considerable attention and succesful results. However, in the last times the higher level fusion procedures (L2, L3) have started to generate a growing interest. These levels aim at obtaining a description of the relations between the objects in the scenario.

These relations are expressed in symbolic terms (actions, intentions, threats), instead of the numerical measures (density functions, movement vectors) computed in L0 and L1.

Context-aware systems aims at the development of computational systems that process data acquired from different datasources and adapt their behaviour in order to provide the 'right' information, at the 'right' time, in the 'right' place, in the 'right' way to the 'right' person (Fischer, 2012). Modern applications must be able to work in problems where the world-behavior is very complex and unpredictable, and where contextual influences are important or even critical. This requires the implementation of flexible and dynamic models, able to adapt to unexpected situations, as well as the exploitation of context knowledge. Cognitive approaches propose building a symbolic model of the world, expressed in a logic-based language, to abstractly represent scene objects and their relations. Cognitive approaches are more robust and extensible than quantitative proposals, but they require the development of suitable interpretation and reasoning procedures, which is not assumable or even possible in all cases. In addition, cognitive models must implement procedures to bridge the gap between abstract representations in the symbolic language and concrete measures acquired by sensors, which is known as the grounding problem. In this sense, ontologies have recently received a considerable attention as proper formalisms to create symbolic models. Ontologies support formal information representation and reasoning while promoting knowledge reuse. These properties make ontologies very suitable in this context, which entails the use of a common communication language between the actors involved in the process, and the integration of several heterogeneous information sources.

The core model, based on the JDL, is stepped in several levels ranging from low-level track data to high-level scene situations. These levels are:

- Tracking Entities (TREN) level, to model input data coming directly from sensors or tracking algorithms: track information (color, position, speed) and time to support the temporal consistency.
- Scene Objects (SCOB) level, to model real-world entities, properties, and relations: moving and static objects, topological relations, etc.
- Activities (ACTV) level, to model behavior descriptions: grouping, approaching, picking an object, and so forth.
- Impact (IMPC) level, to model the association between a cost value and an activity description.
- Process assessment (RECO) leve, to model the feedback process between high-level conclusions and low-level configuration changes.

These abstract ontologies are the building blocks of application-specific knowledge models. Each ontology level provides a skeleton that includes general concepts and relations to describe very general entities and relations, in such a way that they can be extended with more concrete concepts and relations to suit to the requirements of a specific domain. It is interesting to note that the amount of data in lower level ontologies is larger than higher levels. The model has been designed to promote extensibility and modularity. This means that the general structure can be refined to apply this model to a specific domain. Local adaptations should not cause cascade changes in the rest of the structure.

This structure of ontologies may contain both perceptual and contextual data. Perceptual data is automatically extracted by sensors and tracking algorithms, while the context data is external knowledge used to complete the comprehension of the scene. Context data includes but is not limited to information about scene environment, information previously computed and user-requested information. For example, the description of a static object in the scene (size, position, kind of object, etc.) is regarded as context data.

Next subsections indicate general concepts and design principles for each ontologies. These basic information will be widely extended and discussed in the subsequent chapters. More details about the structure of the core model described in this section can be found in (Gómez-Romero et al., 2011b) (Gómez-Romero et al., 2009), (Gómez-Romero et al., 2011b), (Gómez-Romero et al., 2011c) (Gómez Romero et al., 2013).

2.2.1 Tracking data (L1)

The TREN (TRacking ENtities) ontology includes axioms about concepts and relations to symbolically represent raw data directly obtained by sensors or by low-level recognition algorithms.

An ontology design pattern proposed by the W3C Semantic Web Best Practices and Deployment Working Group to define ternary relations in OWL ontologies[1] allow the representation of tracks' temporal evolution, and not only its state in a given instant. In order to keep all the information related to a track during the complete sequence (position, size, velocity, etc.) various set of property values must be associated to each track that are valid only during some timestamps, represented through the OWL-Time `DateTimeDescription`[2]).

Additionally, track properties must be defined as general as possible, in such a way that they can be easily extended. To solve this issue, we have followed the qualia approach, used in the upper ontology DOLCE (Gangemi et al., 2002). This modeling pattern distinguishes between properties themselves and the space in which they take values. This way, we have associated properties to `ActiveTrackSnapshots`, such as `TPosition` or `TSize`. `TPosition` is related with the property `TpositionValue` to a single value of the `TPositionValueSpace`. A `2DPoint` is a kind of `TPositionValueSpace`. The definition of geometrical entities has been developed according to the proposal described in (Maillot et al., 2004), which defines primitive concepts such as *Point*, *PointSet*, *Curve* (as a subclass of *PointSet*), or *Polygon* (a kind of *Curve*).

Additional axioms or rules to calculate complex properties of tracks (e.g. distances), as well as spatial relationships (inclusion, adjacency, etc.), could also be considered and created in TREN.

[1] Defining n-ary relations on the semantic web. W3C semantic web best practices and deployment working group note. `http://www.w3.org/TR/swbp-n-aryRelations/` Last accessed December 2015

[2] Time ontology in OWL. W3C working draft. `http://www.w3.org/TR/owl-time` Last accessed December 2015

2.2.2 Scene objects (L1-L1/L2)

The SCOB (SCene OBjects) ontology includes axioms about concepts and relations to symbolically represent real-world objects and their correspondence with detected tracks.

The root concept in the SCOB ontology is `SceneObject`. Scene objects have properties; e.g. position, illumination, behavior, etc., which may vary in the sequence. To represent properties, we have applied the same combined snapshot/qualia approach as in the TREN ontology. It can be noticed that tracked object property values may be different from the property values of the associated track snapshots, but most of these object property values will be easily inferred from the associated track.

2.2.3 Activities (L2)

The ACTV (ACTiviTies) ontology includes axioms about concepts and relations to describe relations between objects that last in time. This ontology includes axioms involving concepts and relations to describe simple and complex activities. For convenience, these relations have been reified as classes descending from a top concept named `Situation`. In order to establish the temporal duration of the situations, some properties that follows the same pattern based on snapshots described for the lower layers of the model have been introduced.

2.2.4 Impacts and threats (L3)

The IMPC (IMPaCts) ontology has been defined on top of the ACTV ontology. This ontology includes relations to associate situations (instances of the `Situation` concept) and impact evaluations (instances of the IMPC concept `Impact`). This value is a simple numerical assessment or, more probably, a complex expression suggesting or predicting future actions. The qualia approach has been also applied in this ontology.

2.2.5 Process assessment (L4)

Process assessment knowledge includes certain metainformation about the functioning of the framework that is used to improve it. Accordingly, the RECO (RECOmmendations) ontology includes concepts and relations to represent actions that must be carried out to modify either the instances of the ontologies or the behavior of the system. The main concept in RECO is `Action`, which abstractly represent any action that can be understood and carried out by the framework.

3

Video based AmI: A smart home prototype

Ambient Intelligence (AmI) aims at the development of computational systems that process data acquired by sensors embedded in the environment to support users in everyday tasks. Visual sensors, however, have been scarcely used in this kind of applications, even though they provide very valuable information about scene objects: position, speed, color, texture, etc. This chapter shows an adaptation of the cognitive framework presented in 2 for the implementation of AmI applications based on visual sensor networks. The framework combines a priori context knowledge represented with ontologies with real time single camera data to support logic-based high-level local interpretation of the current situation. In addition, the system is able to automatically generate feedback recommendations to adjust data acquisition procedures. Information about recognized situations is eventually collected by a central node to obtain an overall description of the scene and consequently trigger AmI services. The extensions and adaptations of the approach are showed through a prototype system in a smart home scenario.

3.1 Introduction

Ambient Intelligence (AmI) envisions a future information society where users are "proactively, but sensibly" provided with services that support their activities in everyday life (Augusto, 2007). AmI scenarios depict intelligent environments capable of unobtrusively recognize the presence of individuals and seamlessly react to them. To achieve this goal, AmI systems embed a multitude of sensors in the environment that acquire and exploit data in order to generate an adequate response through actuators. Different sensor and network technologies are frequently applied: shortrange (e.g. RFID, NFC); medium-range (e.g. Wi-Fi, Ultrawideband); and large-range (e.g. 4G cell networks). Nevertheless, visual sensors have received less attention, despite the large amount of interesting data that they can obtain from the environment. This gap is mainly due to two reasons: (1) processing visual data is computationally expensive and needs powerful equipment, including a considerable bandwidth to transmit captured images; (2) interpreting visual data is a complex task which may require the use of complex data models, as well as the incorporation of heterogeneous and maybe distributed data sources.

Classical techniques –those strongly based on observational data and a priori knowledge models– have proved to be insufficient to successfully recognize situations in unpredictable and complex scenarios (Gómez-Romero et al., 2010). A solution to overcome these problems has been to provide image-processing algorithms with additional knowledge not directly provided by the cameras; i.e., context knowledge. In video processing, context encompasses any external knowledge used to complete the quantitative data about the scene computed by straightforward image analysis algorithms, including (Bremond and Thonnat, 1996): (1) the scene environment: structures, static objects, lighting and other behavioral characteristics, etc.; (2) the parameters of the recording: camera, image, and location features; (3) stored information: past detected events; (4) soft information provided by human users.

This chapter present an AmI framework that jointly manages visual sensor data and contextual information to support the construction of a symbolic description of the current scene. The knowledge model of the framework is an ontology (Gruber, 1993), which allows representation and reasoning with these types of knowledge. Visual data is firstly processed at single smart cameras to achieve a local interpretation of the situation expressed with instances of the scenario ontology. Different procedures to obtain this local interpretation can be plugged into the framework; in this case, rules involving terms of the ontology are used. The configuration or the behavior of low-level image processing algorithms may be modified according to the local interpretation. Eventually, these interpretations are sent to a coordination agent, which manages the global view of the scene. Information fusion is performed at two levels: (1) heterogeneous data (contextual and sensor based) is used to obtain local scene interpretations; (2) multi-camera information is gathered to build the overall picture of the scenario.

3.2 Proposal

Most research works in the Computer Vision literature have only taken into account local context measures (computed from the pixel values surrounding an object) to accomplish scene recognition (Yilmaz et al., 2006), (Yang et al., 2009). These methods are hardly extensible to different domains, since they often apply application-dependent heuristic calculations. In contrast, cognitive approaches (Vernon, 2008), (Pinza et al., 2008) like the one presented here propose to build a symbolic model of the world expressed in a logic-based language to represent environment objects and relations. Thus, the latter are more extensible, although they require the development of suitable data acquisition and information processing procedures.

The chapter proposal is the implementation of the architecture presented in 2.1 for a video based AmI application and the adaptation of the ontology based cognitive model to data, context and situations from a visual sensor network (VSN) environment. The model allows symbolic manipulation of scene data, in contrast to the classical numerical proposals. Objects and relations-particularly, spatial and topological are abstractly represented. Symbolic representations may lack the precision of numerical approaches, but this is not crucial in most AmI applications. A qualitative representation of the objects' relative positions is enough in most cases to obtain a convenient interpretation of the scenario. Moreover, low-level calibration of the cameras may not be necessary. The model could be populated by other sensor systems in addition to the VSN, as long as they are able to express acquired data in terms of the ontology, thus providing support to multi-modal AmI. Different activity interpretation procedures can be used within the framework as well.

It is important to highlight that the hierarchical architecture implemented in the framework allows task distribution among the cameras while minimizing the amount of exchanged information. This reduces computational requirements and bandwidth consumption of the system.

3.3 Related work

Most works in the literature on general AmI systems tackle the problem of representing and exploiting context information.

Nevertheless, few of them deal with visual information, as previous surveys show (Hong et al., 2009). Currently, there are some promising approaches resulting from the synergies between AmI and the video-surveillance research area a typical application domain of video-based systems, since both of them are concerned with the monitoring of complex environments. Some of these proposals combine multimodel information to achieve scene recognition (Turaga et al., 2011), although they usually apply numerical techniques, which are less flexible and sometimes hardly extensible. Not surprisingly, the need of integrating heterogeneous sensor/context data and the existence of several distributed data sources has resulted in the application of DIF paradigms and techniques to the problem (Steinberg and Rogova, 2008), (Remagnino and Foresti, 2009).

This section focus on research works that apply ontologies to model situations recognized from visual data in AmI applications. One of the most notable contributions is presented in (Snidaro and Foresti, 2007). The authors describe some issues and methodologies for the creation of ontologies supporting AmI applications focused on surveillance and security. Rules are used to create expressions to detect complex events. Some of the main problems that appear in this kind of systems are tackled: event representation, spatial reasoning, uncertainty management, etc. Previously, other approaches –such as PRISMATICA (Velastin et al., 2005)– had also studied the problems that appear in surveillance-related AmI applications, though they are more focused in the management of the VSN, instead of the possible contributions of this technology to AmI.

More recently, ontologies and visual inputs have been combined to detect abnormal behaviors, also in the surveillance domain. The research works in (Vallejo et al., 2009), (Albusac et al., 2010), describe a multi-agent knowledge-based system to characterize and detect abnormal situations in surveillance areas. Interestingly enough, this proposal incorporates imprecise and vague information in the knowledge representation. Multi-agent systems are also used in GerAmi, an AmI environment to supply care and support to elderly people which strongly relies on RFID and Wi-Fi technologies (Corchado et al., 2008).

Insofar as general and high-level ontology-based scene recognition is concerned, an ad hoc proposal for scene interpretation based on DL is presented in (Neumann and Möller, 2008). The paper shows how the reasoning features of the Renamed Abox and Concept Expression Reasoner (RACER) provide functionalities that support scene recognition. The approach is hardly extensible, but it illustrates the expressivity of DL for such tasks, as well as the existence of appropriate tools. DL are also used for modeling and reasoning about complex situations in (Springer and Turhan, 2009). This paper discusses the features required to an ontological model for context-based situation recognition from sensor data, as well as the architectural and implementation details of the corresponding AmI applications. The proposed representation is very similar to the upper levels of our model, but the grounding problem is also solved (Pinza et al., 2008); i.e., the gap between real-world signals and high-level symbolic representations is bridged. This problem is not tackled in (Springer and Turhan, 2009), which does not explain how high-level ontological descriptions are obtained from camera data. Alternatively, probability theories, such as Markov logic networks, have been also used in fusion-based object and scene recognition. The hybrid approach presented in (Wu and Aghajan, 2011) successfully combines low-level image processing and high-level situation description. In contrast, the framework presented emphasizes the role of the cognitive knowledge model, which facilitates reasoning and human interaction, while promoting knowledge reuse. However, we require the creation of proper abduction procedures, which may be difficult in some cases, and it assumes the existence of accurate tracking and identification modules, which is not always possible.

3.4 Data and Information Fusion in visual sensor networks for AmI

A VSN involves the deployment of a certain number of cameras in a wide area –probably with overlapping fields of view– which acquire visual data from the environment. Necessarily, suitable procedures to interpret data captured by single cameras must be developed, in order to obtain an integrated and high-level view of the situation. The DIF research area studies the problems arisen from the combination and interpretation of multiple data sources, at different abstraction levels, and specifically those that appear when data sources are video cameras. As stated in 2.2 the JDL model establishes five operational levels in the transformation of input signals to decision-ready knowledge, namely: signal feature assessment (L0), entity assessment (L1), situation assessment (L2), impact assessment (L3), and process assessment (L4). Below, some of most important problems that appear in VSN-based AmI systems at each JDL level are summarized and the techniques, algorithms, and procedures that are considered in our framework to solve them are briefly presented.

3.4.1 Level 0

3.4.1.1 Camera calibration and data alignment

Information in a VSN must be aligned to a common reference frame. Camera calibration, or common referencing, is the process to calculate the homography matrix that converts from the local coordinates of each camera to a global coordinate space. Calibration can be an off-line procedure (based on the correspondence of the position in the camera plane and in the global plane between of predefined landmarks) or an on-line procedure (based on the analysis of in-use system data; e.g., correspondences between automatically detected corners, edges, etc.).

Numerical calibration of the cameras may not be required in this framework. As explained in 3.5, the smart cameras only interchange high-level descriptions of the perceived situation in terms of topological relations between entities; for example, *a person is close to couch_1*.

If the same identifier *couch_1* is assigned in every camera that is detecting this object, a rough correspondence between their view fields is established. This correspondence may serve as an implicit calibration to align data in the central node. Obviously, this approach is too far to completely solve the problem of calibration, but may be sufficient in several AmI domains where high precision is not required.

3.4.2 Level 1

3.4.2.1 Object detection

The most elemental information that can be extracted from a video sequence is that of the discovery of moving objects. There are various techniques for object detection: (1) temporal differencing, based on the calculation of the pixel-by-pixel difference between consecutive frames; (2) background subtraction, based on the calculation of the difference between the current frame and a predefined background image; (3) statistical methods, based on the difference of additional features extracted from the image; (4) optical flow, based on the computation of the flow vectors of moving objects; and (5) classification, based on the identification of a pattern in the image with trained classifiers.

Object detection is not trivial, since in most cases the conditions of the watched environment change. For example, changes in the lighting and the shadows of the objects during daytime (especially in outdoors applications) and moving objects that become static must be taken into account. Object detection is performed in the framework by the tracking layer (see 3.5), which relies on a tracking procedure. The framework can be configured to use different tracking algorithms.

3.4.2.2 Object tracking

Detected objects must be tracked over time; i.e., the system must segment the moving objects and assign consistent labels during their complete lifecycle. Specifically, a track is defined as a set of groups of connected pixels that represent a moving object with some properties: size, color, speed, etc. In the simplest case, a track includes a single group of connected pixels. Tracking is defined as the estimation of the number of objects in a continuous scene, together with these properties: locations, kinematic states, etc. Object tracking has been tackled by applying statistical prediction and inference methods, such as Kalman or particle filters, adapted to visual data association. The tracking layer of the framework performs the complete tracking procedure, as explained in 3.5.

Estimation techniques are very sensitive to the particular conditions of the scenario, and therefore they may be insufficient in some applications. The incorporation of context knowledge has been regarded as essential to deal with complex scenarios with occlusions, lighting changes, and object deformations. In our framework, object and situation information at levels 2 and 3 (obtained by applying context knowledge) is used to change the parameters of the tracker according to the current scenario and past events.

3.4.3 Level 2

3.4.3.1 Classification

Object identification and activity recognition are two fundamental classification tasks that must be performed in an AmI application (VSN-based or not). Object identification aims at determining the type of a tracked object; e.g., person, bottle, etc. Thus, it can be considered halfway between L1 and L2. In the framework, a priori rules are used to classify objects according to their features –mainly the size (see 3.7.2). This approach should be extended with more advanced techniques and/or machine learning enhancements, in order to automatically classify tracks according to other features: color, histogram, etc.

Activity recognition, in turn, aims at discerning that an activity is taking place. Usually, two types of activities are distinguished: basic activities –i.e. simple activities that cannot be decomposed into more simple actions (e.g., walking), and composite activities –i.e., activities that are the result of various simple actions (e.g., laying the table).

Activity recognition is an open problem in general applications, since it requires systems to develop cognitive capabilities close to human understanding. In this chapter, pre-defined rules are used to identify activities from moving object properties and context information. The main strength of rules is that we can express almost any condition by using terms defined in the ontological model. In contrast, at its current state, they must be manually created, which requires a considerable effort. In addition, other methods could be applied in the framework in combination to rules to identify complex activities. The use of the symbolic models facilitates the integration of these different techniques, since any procedure can be plugged into the framework as long as its output (i.e., recognized activities) is expressed with the same ontology language.

3.4.4 Level 3

3.4.4.1 Situation assessment

Level 3 focuses on the estimation of the impact of a situation of the application domain. In other words, situation assessment is the process of detecting and evaluating particular situations that are of special relevance to the scenario because they relate to some type of threatening, critical situation or any other special world state. This JDL level includes procedures for the identification of abnormal and hazardous situations, which is especially relevant in some AmI domains; for example, Ambient Assisted Living applications require implementing proper mechanisms to react to an emergency situation if the user does not follow the normal sequence of activities, it falls down, or it abruptly interrupts an ongoing activity. The framework applies the same rule-based mechanism explained for Level 2 tasks: rules with terms of the lower abstraction level are used to create instances representing information at this level.

3.4.5 Level 4

3.4.5.1 Process enhancement

Process enhancement –also known as active fusion– is aimed at the modification of the data acquisition and processing procedures after DIF to enhance results quality. Generally speaking, process enhancement consists in improving a fusion procedure by using feedback generated at a more abstract level. For instance, the behavior of a tracking algorithm can be changed once a general interpretation of the scene has been inferred; if the system recognizes that an object is moving out of the camera range through a door, the tracking procedure could be informed to be ready to delete this track in the near future. As previously mentioned, the framework includes a general mechanism to generate recommendations for the tracking procedure based on rule triggering. In their basic form, these recommendations are direct manipulations of the parameters or the data stored by the tracker, as it is exemplify in 3.7.3.

3.5 Framework architecture for AmI applications

The architecture of the VSN framework is depicted in Fig. 3.1. The schema shows the two types of nodes that are defined in the VSN: the smart cameras and the central node.

Smart cameras are video cameras able to perform DIF tasks. Cameras capture data, which is processed by a low level tracker. Tracking information is then introduced into the abstract scene model as ontology instances. Rules are activated as a result of the changes of scene objects detected by the tracker. Eventually, as a result of the model-building process, these rules create instances corresponding to situations. Situations detected by single smart cameras are sent to the central node. The contents of the messages between smart cameras and the fusion node are encoded with the same ontology used for the smart camera scene model. The central node processes the situations detected by the cameras in order to obtain a more complete view of the scene.

Smart cameras process data at two logical levels: (1) the tracking layer; (2) the cognitive layer. First, each camera is associated with a process that acquires video frames. Next, the tracking sub-system sequentially executes various image-processing algorithms to detect and trace all the targets within the local field of view. The tracking layer is arranged in a pipelined structure of several modules, which correspond to the successive stages of the tracking process (Besada et al., 2005), (Patricio et al., 2007): (1) detection of moving objects; (2) region-totrack multi-assignment; (3) track initialization/deletion; (4) trajectory analysis.

Tracking data is introduced into the cognitive layer to initiate more complex high-level information fusion procedures. Smart cameras implement an a posteriori schema for context information exploitation (Gómez-Romero et al., 2010). This schema proposes the implementation of a processing layer on top of the tracking procedure. In this layer, abstract ontologies are used to describe abstract entities. The tracking layer and the cognitive layer communicate through an interface, which offers methods to revise the ontological model in the update and initialization/deletion steps. In the next section, the structure of the ontologies and the processes to create ontology instances in the cognitive layer is described.

Communication between the smart cameras and the central node is started when a new situation is detected –i.e., when a new instance of the concept that represents a situation is created in the symbolic model (see 3.6). The detected situation is sent to the central node, expressed in the suitable situation ontology. The use of a formal ontology to communicate situation information facilitates the incorporation of heterogeneous cameras –or even other sensors– to the system, as long as they are able to use the same situation ontology to communicate information.

The central node gathers camera information to build a unified view of the scene. This unified view is represented with instances of the same ontologies used for smart cameras. The combination of local camera information has been implemented with a rule-based mechanism, as explained in 3.8.

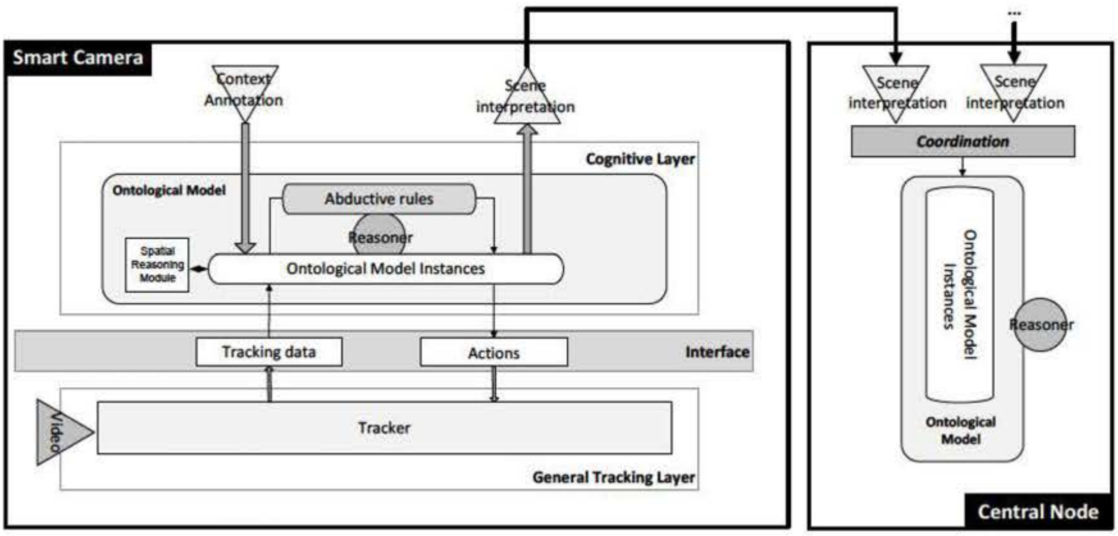


Figure 3.1: Architecture of the AmI framework: Smart Cameras and Central Node

3.5.1 Model construction

Scene interpretation consists in obtaining a symbolic model of the scene activities. Ontologies support the definition of a formal vocabulary to create these symbolic models. This vocabulary includes the terminological axioms (i.e., axioms about classes and relations) that are used to delimit the possible realizations of the model. In DL nomenclature, the set of axioms defining concepts is the TBox of the ontology, whereas the set of axioms defining properties is the RBox of the ontology. The concept and relation instances of the ontology are defined with axioms about individuals, which represent the evolution in time of the scene tracks, objects, situations, etc. These axioms about instances of the ontologies compose the ABox. Essentially, scene interpretation is a model-building procedure in which instances of the concepts and relations defined in the scene vocabulary are created.

3.6 Knowledge representation and reasoning

The ontological model of the cognitive layer encodes the context information provided, the perceptions acquired by cameras, and the model obtained after reasoning processes. To manage these three types of scene knowledge, a set of layered interrelated ontologies organized according to the abstraction layers defined in 2.2 is proposed.

The overall structure of the VSN ontologies is depicted in Fig. 3.2. It is important to take into account the differences between the general knowledge (i.e., very abstract knowledge that is common to any context-aware application) and the specific knowledge (i.e., knowledge specific to a concrete AmI application). Accordingly, the upper ontologies that contain terminological axioms defining basic concepts and relations are TREN, SCOB, ACTV, IMPC and RECO.

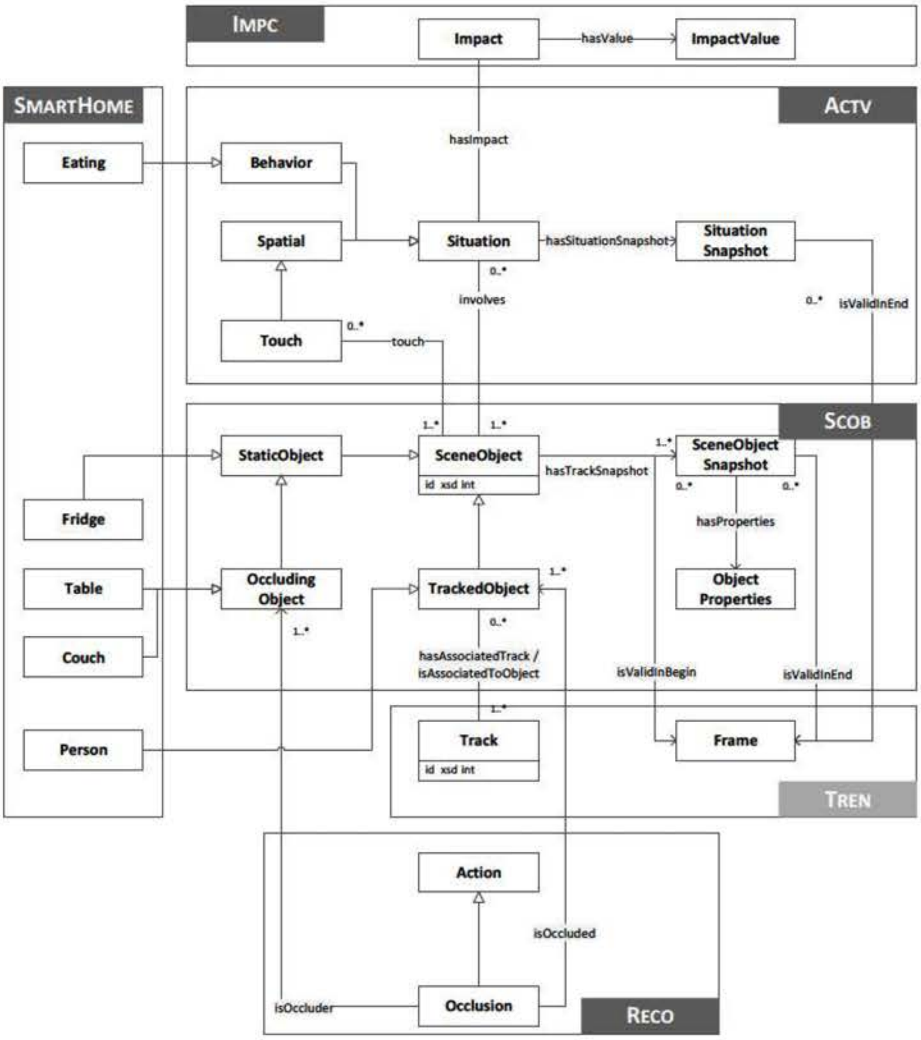


Figure 3.2: UML excerpt of the high-level ontologies: main concepts and grounding

The concepts and relations defined in these upper ontologies must be specialized in each application. As can be shown in figure 3.2, the new ontology SMARTHOME has been defined by refining the concepts included in the general ontologies. The purpose of this separation is to provide final application developers with a general knowledge frame with well-defined building blocks, in such a way that they only need to extend it to model new scenarios. For example, the SCOB ontology defines the `OccludingObject` class, which –for the sake of simplicity– is a type of `StaticObject`. In SMARTHOME, the concept `Couch` is defined as a subclass of `OccludingObject`, and consequently it inherits all its properties.

It is interesting to note that these ontologies are closely related between them. In fact, they represent the transformation from low-level tracking data to high-level situation knowledge. An ontology of an upper abstraction level is linked (or grounded) to an ontology of a lower abstraction level. Accordingly, the ontology for scene objects defines the property `hasAssociatedTrack` to associate instances of scene objects to instances corresponding to track data. Thus, information at object level is described in terms of objects and objects' relations, but objects are associated to the tracks obtained by the tracking layer. Similarly, a more abstract ontology is defined to represent scene situations; these situations are grounded to the involved objects represented in the scene objects ontology.

Contextual knowledge is introduced into the model as instances of the proper ontologies, which is known as scenario annotation. Annotations include object position and size, possible occlusions, enter and exit zones, or any other convenient contextual knowledge. This zeropoint knowledge is used in the reasoning process that is activated when moving objects are detected in the scene.

Additionally, reasoning rules (deductive or abductive) are introduced into the knowledge model (see 3.7). The combined use of ontology specialization and rules allows the definition of very general rules that are triggered with objects of the classes and the subclasses. For instance, a general rule to detect proximity between a `TrackedObject` and an `OccludingObject` will be fired not only with direct instances of these concepts, but also with instances of their subclass; e.g., `Person` and `Couch`, respectively. In this way, a new entity can be described as a subclass of many existing classes, and consequently it defines its behavior as the composition of the behavior of its superclasses.

In the remainder of this section some details about the structure of the terminological part of the ontologies provided with the framework are explained and the nature and the implementation of reasoning procedures within the representation model are discussed.

3.6.1 Adaptation of knowledge representation

3.6.1.1 Tracking data (L1)

Instances of the TREN ontology are created as a result of the initialization and the update stages of the low level tracking algorithms (see Fig. 3.3). The core concepts in TREN are `Frame` and `Track`. A `Frame` is identified by a numerical ID and is marked with a timestamp using an OWL-Time `DateTimeDescription`, this allow the association of a set of `Snapshots` to each `Track`. Each `Snapshot`, representing track feature values, is asserted to be valid in various `Frames`.

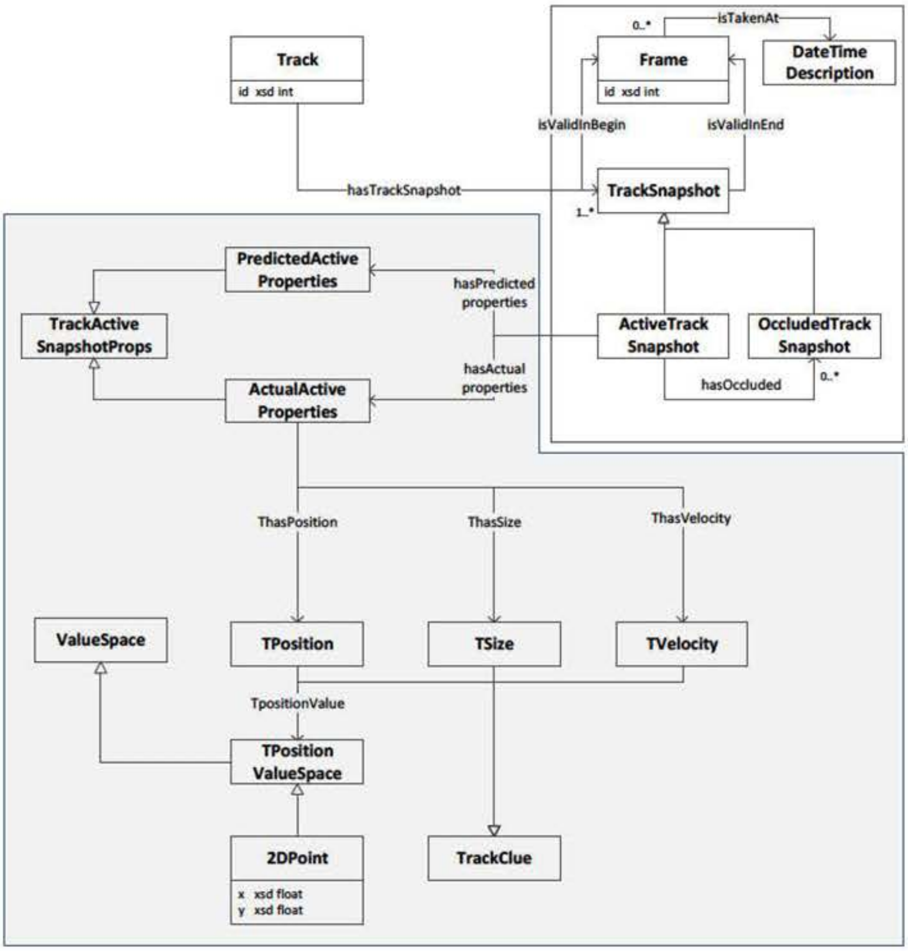


Figure 3.3: UML excerpt of the TREN ontology: representation of track properties and track snapshots

3.6.1.2 Scene objects (L1-L1/L2)

The SCOB ontology includes both static and dynamic objects. Static objects (class `StaticObject`) are scene objects defined a priori. Not surprisingly, most of contextual entities are instances of the `StaticObject` class. Dynamic objects (class `TrackedObject`) are scene objects detected during the functioning of the framework. Instances of dynamic objects are created as a result of correspondence and reasoning procedures. `StaticObject` and `TrackedObject` are subclasses of `SceneObject`, the root concept in the SCOB ontology.

3.6.1.3 Activities (L2)

As mentioned in 3.4, this ontology is used for communication between smart cameras and the fusion node of the architecture. Simple activities, expressed as instances of the ACTV ontology, are sent to the central node to be combined with other inferences to eventually detect a complex situation. Complex situations are also introduced as instances of ACTV, in this case in the local instantiation of the ontology managed by the central node.

3.6.1.4 Impacts and threats (L3)

Impact and threats are the most application-dependent knowledge of the ontological model; therefore, they must be conveniently specialized in a given domain. In this case, as the application example does not require impact evaluations no further adaptations or extensions are required.

3.6.1.5 Process assessment (L4)

In the example application in 3.7 these recommendations are instances generated as a result of rule triggering. Once a recommendation is created, it is synchronously executed, since delaying the modification may be unproductive or error-prone. A more complex policy for handling recommendations could be developed; as a matter of fact, the basic mechanism could be extended to implement a priority queue to asynchronously retrieve, interpret, and carry out the procedures specified by `Action` instances.

3.6.2 Deductive and abductive reasoning

Standard ontology reasoning procedures are performed within the ontologies to infer additional knowledge from the explicitly asserted facts. To name some of them, the inference engine supports tasks such as classification (i.e., to determine the class hierarchy of an ontology) and instance checking (i.e., to determine the classes which an instance belongs to).

Ontology Web Language standard does not directly support deductive rules, but several extensions have been proposed. One of the most extended is SWRL (Semantic Web Rule Language) (Horrocks and Patel-Schneider, 2004), which allows deductive inference within OWL ontologies. Rule-based formalisms can be used with limitations, since reasoning with models combining rules and OWL is decidable only under certain safety restrictions (Motik et al., 2005). Deductive rules are used to maintain the consistency of the ontology and to explicitly assert axioms involving existing instances. An example of deductive rule would be "the position of a `Track` must be the same as the position of the last associated `TrackSnapshot`".

Monotonicity of ontology languages forbids adding new knowledge to the models while reasoning, which is required in scene interpretation. Actually, scene interpretation is a paradigmatic case of abductive reasoning, in contrast to the DL deductive reasoning (Elsenbroich et al., 2006). Abductive reasoning is defined as a form of reasoning that takes a set of facts as input and draws a suitable hypothesis that explains them –sometimes with an associated degree of confidence or probability. This type of reasoning is also called Inference to the Best Explanation (IBE). Visual data interpretation can be regarded as an IBE process where the objective is to figure out what is happening in the scene from the observed and the contextual facts. In terms of the knowledge model presented in the previous section, scene interpretation can be seen as an abductive process to generate ontology individuals of a higher-level ontology from instances of a lower level ontology.

Abduction is not directly supported by DL ontologies, but it is simulated in some reasoning engines by defining non-standard inference rules. These rules allow the creation of new instances in the consequent, which is forbidden in standard rules to satisfy the safety condition. The RACER (Haarslev and Möller, 2001) engine, which is used in this work, supports abductive reasoning through extension rules that create or modify instances of the ontology representing scene interpretations. Please note that these rules do not directly support representation of uncertain knowledge. As shown in the next section, uncertainty management may not be essential in simple AmI scenarios, but it must be considered in more complex domains involving scene recognition (García et al., 2011) (see 5)

This framework allows two types of non-standard rules: bottom-up rules and top-down rules. Bottom-up rules are used in scene interpretation, and as mentioned, they obtain instances of upper-level ontologies from instances of lower-level ontologies. For instance, some rules could be defined to identify objects from track measures; i.e., to obtain instances of the scene objects ontology from instances of the tracking data ontology. An example of simple rule is “create a `Person` instance when an unidentified `Track` larger than a predefined size is detected inside a region of the image”.

Top-down rules are used to create instances of the `Action` concept in RECO from the current interpretation of the scene, the historical data, and the predictions. Topdown rules may result in corrections to the low-level fusion procedure: tracking parameters, data structures, etc. A simple rule would recommend “to ignore a `Track` associated to a `Person` which is inside an area previously annotated as a mirror”.

3.6.3 Spatial reasoning

One key aspect of our model is representation and reasoning with qualitative spatial properties. Ontologies do not directly support spatial reasoning, which has given rise to the development of joint approaches that incorporate suitable additional constructors (Katz and Grau, 2005b), (Grütter et al., 2008) (see B).

RACER includes support for RCC through the activation of an extended reasoning layer, namely the RCC substrate, which allows the use of RCC predicates in representations and queries while preserving RCC semantics. In addition, user-defined relations can be extended with RCC semantics; in the simplest case, this means to make a user-defined relation equivalent to a RCC predicate.

Region Connection Calculus predicates (and RCC equivalent user-defined properties) must be instantiated in the knowledge base. This implies the creation of an instance of a RCC relation between two instances to represent that the corresponding scene entities are disconnected, partially overlapping, etc. To calculate the instantiation of RCC properties, calculations must be performed to obtain the distance between the bounding boxes of two tracks (or objects). This can be alternatively achieved: (a) by using supported lambda calculus expressions to be executed by RACER (Gómez-Romero et al., 2009); (b) by performing a topological analysis in a pre-processing step (Serrano et al., 2011). Our experiences prove that the second approach is more appropriate; otherwise the performance of the reasoning process is seriously compromised, because RCC properties of moving objects must be often recalculated. Additionally, pre-processing facilitates the implementation of additional optimizations and the use of third-party tools supporting topological calculations.

The framework includes a pre-processing module to instantiate RCC properties. This module is executed when a new contextual object is annotated in the scenario (infrequent) or when a tracked object changes its position (very frequent). The module is based on the OpenGIS Simple Features standard, which is a specification for storage of geographical, spatial, and non-spatial attributes and operators[1]. OpenGIS is implemented in the programming interface Java Topology Suite (JTS)[2]. RCC and OpenGIS are not directly compatible, but translations between both specifications can be easily carried out.

Additional improvements could be implemented in the pre-processing module to increase the computation speed. It is interesting to highlight that checking object spatial relations, and particularly RCC relations, has a complexity O(n2) –the test must be performed between each pair of elements. Thus, it would be convenient to build a data structure able to maintain a hierarchical spatial partition on the Euclidean space. Tree structures, such as R-Tree, R*, and quad-trees can be applied, though it must be taken into account that applications with large number of dynamic objects and frequent updates will require very often tree rebuilding. Currently, our framework does not support these improvements, which remains as a promising line for future work.

3.7 Application example

In this section, an example of the use of our framework in a smart home application is provided. Firstly, depicting how the knowledge model is adapted to the scenario; i.e., the creation of contextual rules and ontology instances. Secondly, describing the reasoning procedures performed by the framework: object identification, tracking enhancement, and single camera scene identification. Video sequences included in the LACE dataset of the University of Rochester [3] are used to illustrate this example. This dataset includes footage taken from several cameras covering a room that reconstructs the living room and the kitchen of a studio. Only one moving person is present in the videos. For the sake of simplicity, the output of three cameras located in the room is used, as depicted in Fig. 3.4, which have considerable overlapping fields of view.

The framework allows interoperation between the General Tracking Layer and the Cognitive Layer through the implementation of a Java interface based on ViPER-GT (Video Performance Evaluation Resource-Ground Truth authoring tool) (Doermann and Mihalcik, 2000), (Serrano et al., 2010) and OWLAPI 2 (Horridge and Bechhofer, 2009). The interface stores the ontological model, facilitates scenario annotation, communicates with the low-level

[1] OpenGIS implementation specification for geographic information–simple feature access. http://www.opengeospatial.org/standards/sfa/ Last accessed December 2015

[2] Java Topology Suite web page. http://www.vividsolutions.com/jts/JTSHome.htm Last December 2015

[3] LACE dataset. http://www.cs.rochester.edu/*spark/muri/ Last accessed December 2015 (Not available)

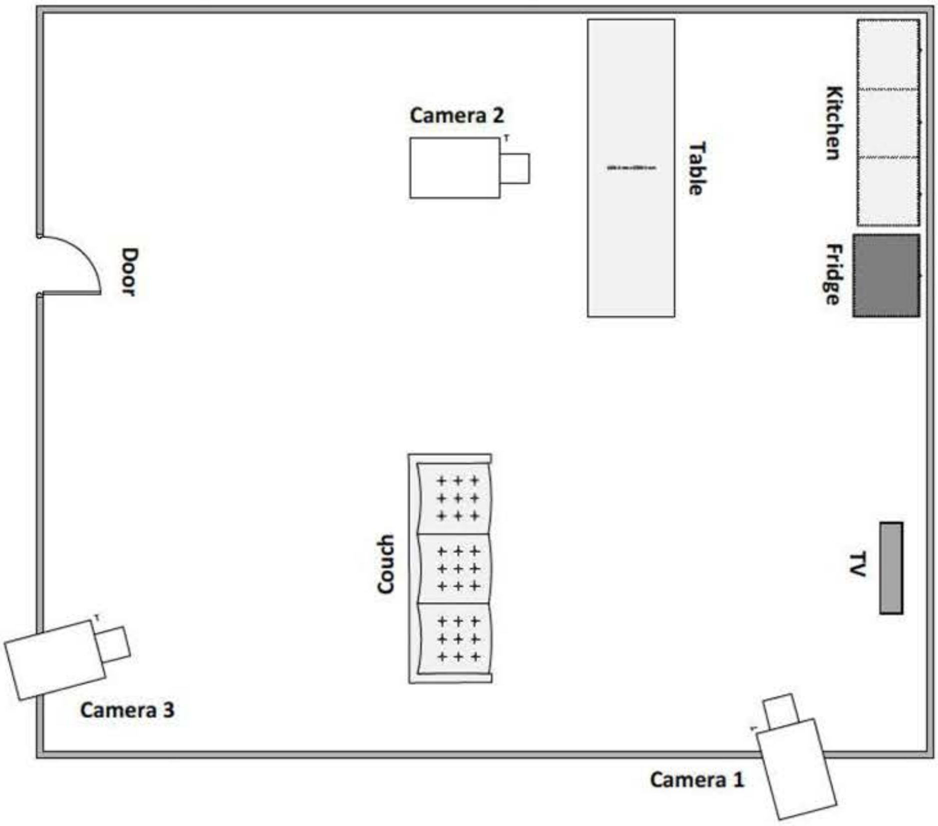


Figure 3.4: Scenario plane: camera and static objects location

tracking procedure(s), interacts with the RACER reasoner to perform inference tasks, and graphically presents the information generated by the framework (tracks, scene objects, etc.) along with the video sequence. In combination with RacerPorter (a graphical user interface to RACER), the software allows the operator to check the results provided by the tracking procedure and the outcomes of the fusion process. The system has been tested with the trackers presented in (Patricio et al., 2007), (García et al., 2005). Notice that each camera runs an instance of the software and has a different context model.

3.7.1 Scenario annotation

Before starting the processing, the framework must be configured; particularly, the scenario viewed by each camera must be annotated. The application-specific ontology, namely SMARTHOME (see 3.6), is extended from general ontologies of the framework. Among others, SMARTHOME includes new concepts for situations and objects:

- Concepts:

  Objects: `Person`, `Door`, `Couch`, `Table`, `Fridge`

  Scenes: `Eating`, `UsingFridge`

- Axioms:

`Person` $\sqsubseteq$ `TrackedObject` (a person is a tracked object)

`Table` $\sqsubseteq$ `OccludingObject` (a table is an occluding object)

`Couch` $\sqsubseteq$ `OccludingObject` (a couch is an occluding object)

`Fridge` $\sqsubseteq$ `StaticObject` (a fridge is a static object)

Figure 3.5 shows the use of the annotation tool to create the context object instances that are initially inserted into the ontology. The same objects are marked in cameras 1, 2 and 3: exit door, couch, table and fridge. The tool automatically inserts proper instances of the respective concepts in the ontology, and assigns property values –mainly, position points. The figure depicts the correspondence between ontology instances and scenario information. The excerpt of the OWL code corresponds to the definition of `fridge1` as an instance of the `Fridge` class with a point of the bounding polygon at position (687, 144).

This procedure must be repeated to initialize the context model of each camera. It is interesting to highlight that the same identifier is assigned to an object regardless of the camera scenario that is being annotated (Fig. 3.6). For example, the fridge has the object identifier 1 in camera 1 and camera 2.

3.7.2 Context-based object identification

After initialization, the SMARTHOME ontology (with corresponding instances) is loaded into the RACER reasoning engine running on the contextual layer of a smart camera. Contextual rules (abductive and deductive) are also introduced into the reasoning engine in this step. These rules can be either general reused rules from the proposed top level ontologies, or particular rules only applicable to the field of view of the camera.

An example rule that introduced in the reasoning engine is: "if a `Track` is bigger than a predefined size, then it corresponds to a `Person`" (Fig. 3.7). This rule is used to identify people appearing on the camera view. The syntax of the rule, is expressed in the Lisp-based nRQL language of RACER. The rule has been created in the context model managed by camera 3 to create a new person instance when a non-identified track larger than (20 x 50) is detected. This value could be stored in the ontology itself as a property of the camera. To do so, the rule checks if there is a track not associated to an object that is currently valid and whose size properties are appropriate. (Notice that terms preceded with `?` are variables that are bound to instances of the ontologies.)

The rule is triggered in frame 45 (Fig. 3.8) consequently, a new instance of the `Person` class is created. The name of this instance is automatically generated by RACER from the provided prefix `(person-ins)` and the suffix `(?t)`. The property `hasAssociatedTrack` is assigned to the new instance to point to the track that has caused the rule firing, and the previous association is removed from the knowledge base. The formulation of the rule shows that retrieving the property values of the current snapshot is not trivial as a result of the qualia approach used to represent generic relations. Nevertheless, RACER offers the possibility of defining stored queries to be re-used in subsequent rules or queries. Therefore, a convenient stored query has been created to retrieve the properties of the current snapshot of a given track.

```
<!-- fridge1 instance -->
<owl:Thing rdf:about="#fridge1">
  <rdf:type rdf:resource="#Fridge"/>
  <scob:hasObjectSnapshot rdf:resource="#osn_fridge1"/>
</owl:Thing>

<!-- object snapshot of fridge1 -->
<owl:Thing rdf:about="#osn_fridge1">
  <rdf:type rdf:resource="&scob;SceneObjectSnapshot"/>
  <scob:hasObjectProperties rdf:resource="#fridge1_props"/>
  <tren:isValidInEnd rdf:resource="&tren;unknown_frame"/>
</owl:Thing>

<!-- properties of fridge1 snapshot (position) -->
<owl:Thing rdf:about="#fridge1_props">
  <rdf:type rdf:resource="&scob;ObjectSnapshotProperties"/>
  <scob:OhasPosition rdf:resource="#fridge1_position"/>
</owl:Thing>

<!-- fridge1 position -->
<owl:Thing rdf:about="#fridge1-position">
  <rdf:type rdf:resource="&scob;OPosition"/>
  <scob:OpositionValue rdf:resource="#p1"/>
  <scob:OpositionValue rdf:resource="#p2"/>
  <scob:OpositionValue rdf:resource="#p3"/>
  <scob:OpositionValue rdf:resource="#p4"/>
  <scob:OpositionValue rdf:resource="#p5"/>
  <scob:OpositionValue rdf:resource="#p6"/>
</owl:Thing>

<!-- fridge1 point1 coordinates -->
<owl:Thing rdf:about="#p1">
  <rdf:type rdf:resource="&tren;2DPoint"/>
  <tren:y rdf:datatype="&xsd;float">687.0</tren:y>
  <tren:x rdf:datatype="&xsd;float">144.0</tren:x>
</owl:Thing>
```

Figure 3.5: Camera 1: contextual objects annotation

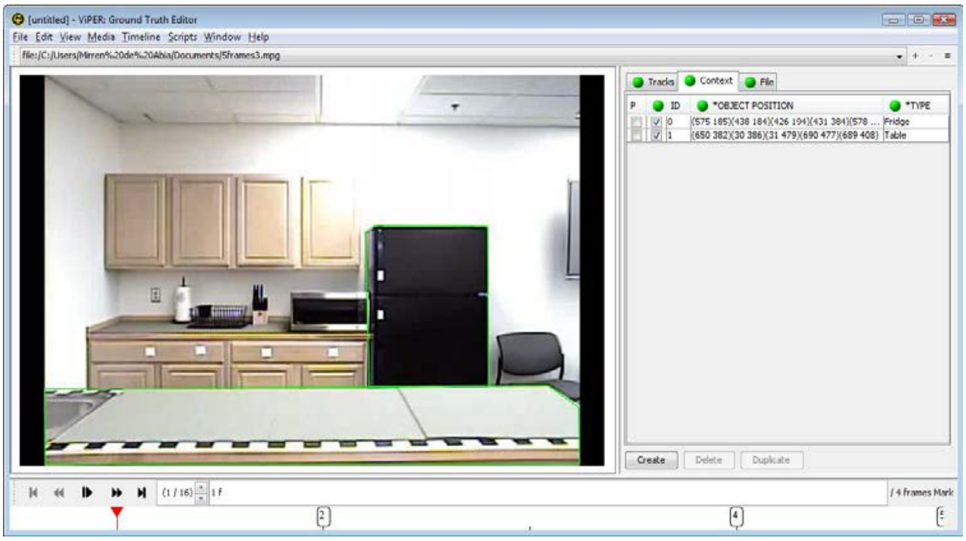


Figure 3.6: Camera 2: contextual objects annotation

```
;;; Correspondence rule [1]
(firerule
       (and
         (?t     #!scob:unknown_object #!scob:isAssociatedToObject)
         (?t     ?tsn                  #!tren:hasSnapshot)
         (?tsn  #!tren:unknown_frame  #!tren:isValidInEnd)
         (?tsn  ?tsnp                 #!tren:hasActualProperties)
         (?tsnp ?tpos                 #!tren:ThasPosition)
         (?tpos ?p                    #!tren:TsizeValue)
         (?p     (>= #!tren:w 20))
         (?p    (>= #!tren:h 50)))
    (
       (instance
         (new-ind person-ins ?t)      #!smarthome:Person)
         (forget-role-assertion
               ?t #!scob:unknown_object #!scob:isAssociatedToObject)
         (related
              (individual person-ins)
               ?t #!scob:hasAssociatedTrack)))
     )
)
```

Figure 3.7: Correspondence rule

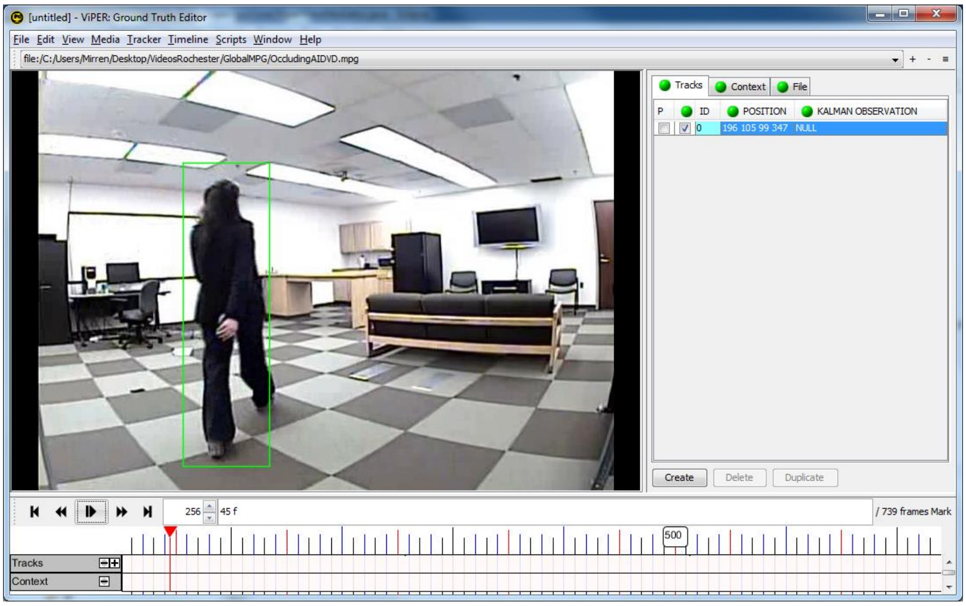


Figure 3.8: Camera 1: correspondence rule is fired

3.7.3 Tracking enhancement

Rules have been as well defined to create actions to enhance the functioning of the low-level tracking procedure. A typical case is finding that a tracked object is overlapping with an occlusive object, in order to predict that it will be only partially detected (or even not detected) in the next frames. As a matter of example, in this section a rule that detects that a person track is overlapping with an occluding object is presented (Fig. 3.9). This rule creates an instance of the `Action` class stating that an occlusion situation has started. If the tracker detects a dramatic change of the size of the track involved in the overlapping situation between consecutive frames while the occlusion is active, it is recommended to keep the previous size of the track.

```
;;; Occlusion detection rule [2]
(firerule
        (and
          (?track        #!tren:Track)
          (?person       #!smarthome:Person)
          (?person      ?track          #!scob:hasAssociatedTrack)
          (and          (?object           #!scob:OccludingObject)
                        (?*track        ?*object         :po)))
   (
        (instance
          (new-ind ind)             #!reco:Occlusion)
          (related (individual ind) ?object    #!reco:isOccluder)
          (related (individual ind) ?person    #!reco:isOccluded))
    )
)
```

Figure 3.9: Occlusion detection rule

This rule combines context knowledge, dynamic knowledge, and RCC-based reasoning (with the 'partially overlaps' `PO` predicate). It is assumed that the: `po` predicate is instantiated as a result of the spatial reasoning performed by the pre-processing. It can be seen that variables involved in RCC predicates must be noted with a special symbol `(?*)`[4].

The rule is triggered in frame 118 of the sequence, being `person1` and `couch1` the objects that match the rule antecedent. (Notice that `couch1` is not a direct instance of `OcluddingObject`, but an instance of the `Couch` subclass.) Therefore, an instance of the `Occlusion` class is created (Fig. 3.10).

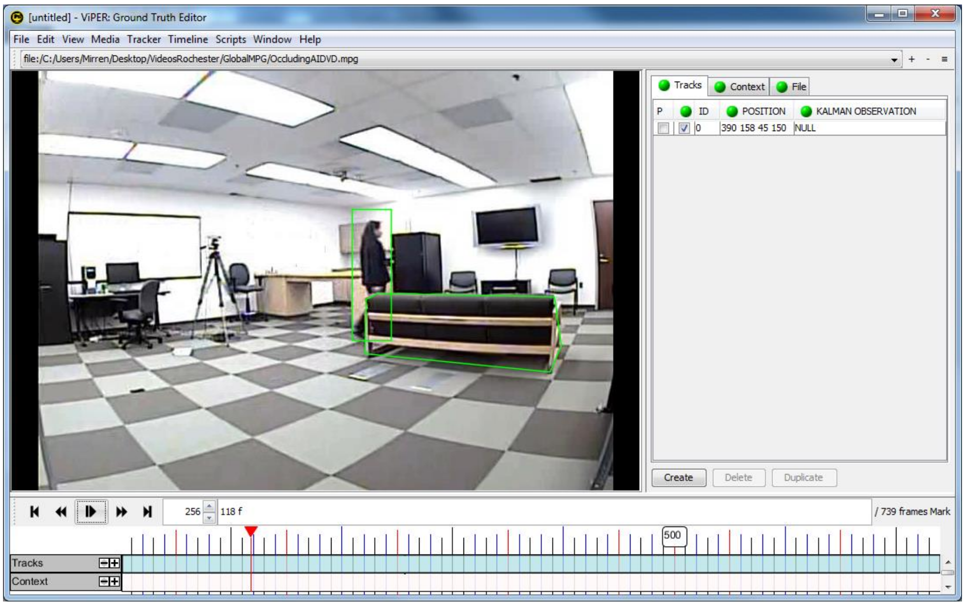


Figure 3.10: Camera 3: occlusion detected

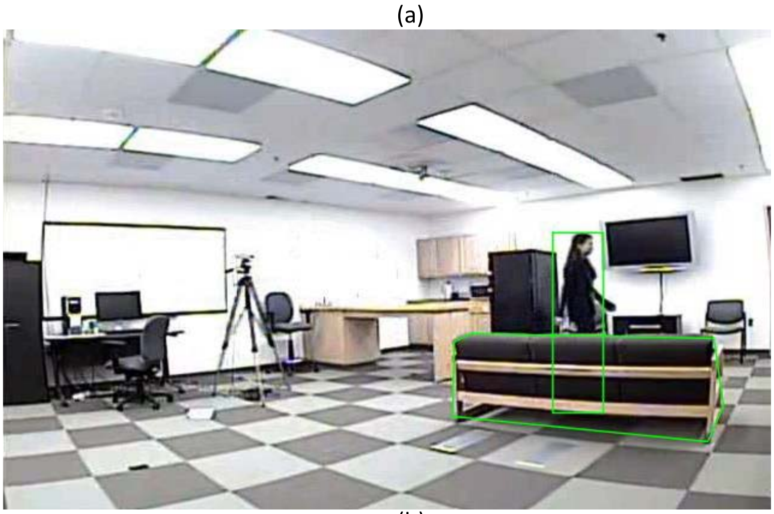
After the new `Occlusion` instance is created, and while the situation is not finished, the framework watches the changes in the size of the occluded object in order to keep the consistency and avoid the effects due to the occlusion. In this example, the procedure is configured to reassign the size and the position of the track to the previous observation when a size change over 80detected. Figure 3.11 shows the bounding box of the track as calculated by the tracker without context and the bounding box as estimated by the cognitive layer as a result of the reasoning procedure.

Figure 3.12 shows a comparison between the positions of the person as detected by the tracking procedure and the positions as recalculated by the cognitive layer during the occlusion –the ground truth has been manually obtained. It can be seen that the use of the context layer considerably improves the results of the tracker.

The Root Mean Square Error (RMSE) of the track size obtained by the general tracking layer and the cognitive layer are, respectively, 940.4 and 486.6 (Fig. 3.13). The modification applied by the cognitive layer is quite conservative, which is correct in this sequence since the changes in the person are not very significant. The graphs show that if actual position changes

[4]Additional information to represent when the recommendation has been created and the starting and ending frames should be added. For the sake of simplicity, this information is omitted.

(a)



(b)

Figure 3.11: Camera 3: occlusion correction. (a) Tracker output, (b) Cognitive layer output

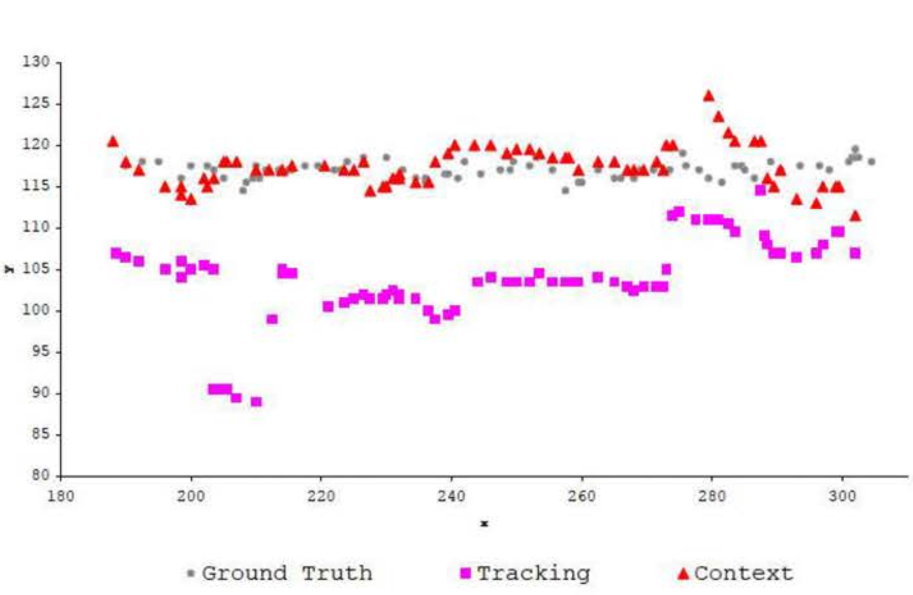


Figure 3.12: Position (x, y) - 'Tracking' vs. 'Tracking+Context'

occurs (e.g., the person falls behind the couch), this policy will lead to errors. Nevertheless, additional rules can be easily created to model these situations and react conveniently.

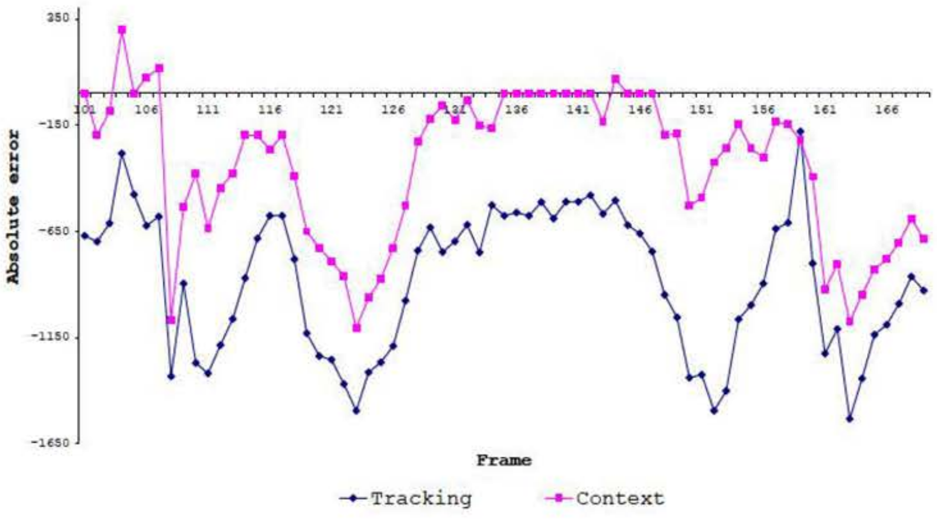


Figure 3.13: Size - 'Tracking' vs. 'Tracking+Context'

The context model includes a similar rule to detect the end of the occlusion. Conversely, the rule for the end of occlusion uses the RCC predicate DC (disconnected). The occlusion is finished, which means that the valid period is closed. In terms of the ontological model, that means assigning a frame other than *unknown frame* to the `isValidEnd` property of the situation. Subsequently, the framework stops watching the size of the track involved in the occlusion. In addition, it must be taken into account that the occlusion detection rule will be also triggered when the person is in front of the couch. Nevertheless, in this case the tracker does not detect any noticeable change in the track size, and therefore the track size is not corrected. The creation of a false occlusion instance, as well as other problems resulting from the 2-dimension information managed by local cameras, is avoided by using the information of more than one camera, as described in 3.8.

3.7.4 Single-camera simple scene recognition

Additional rules have been defined in the model to interpret what is happening in the scene from tracking and object data acquired by a single camera. Our framework focuses on discovering RCC-based spatial relations between annotated objects. These simple situations are represented in the model as instances of the `Situation` class in the ACTV ontology. Therefore, single-camera rules for scene interpretation include object conditions in the antecedent and instructions for ACTV instances creation in the consequent.

For example, a rule (Fig. 3.14) defined in camera 2 detect if a person is enclosed into the fridge object (RCC `NTPP` predicate) –that means that the person is operating the fridge. If the rule is triggered, a new instance of the `Enclosing` situation (defined in the SMARTHOME ontology) is created, as well as a relation between the involved objects via the `enclosed` and `enclosing` properties. This rule is fired in camera 2 at frame 39 of the test video (Fig. 3.15). At this point of the execution, a new `Situation` instance is created in the knowledge model of the camera.

```
;;; Single-camera simple scene recognition (camera 2) [3]
(firerule
       (and
         (?track        #!tren:Track)
         (?person       #!smarthome:Person)
         (?person     ?track          #!scob:hasAssociatedTrack)
         (and         (?object        #!smarthome:Fridge)
                      (?*object       ?*track     :ntpp)))
   (
       (instance
         (new-ind ind ?person ?object)     #!smarthome:Enclosing)
         (related (individual ind) ?person #!smarthome:enclosed)
         (related (individual ind) ?object #!smarthome:enclosing))
    )
)
```

Figure 3.14: Simple scene recognition enclosing rule

Additional AmI services may be launched as a result of this situation; for example, if there is an unsafe equipment instead of the fridge as the touched object, a warning could be launched to the person or to the remote operator. The situation is finished when the termination rule is fired. This second rule also uses the RCC relation `DC` to detect that the person is no longer overlapping with the fridge.

Besides, the new situation information –i.e., the new instances of the `Action` class and other related instances– is sent as soon as detected to the central node. This knowledge is processed and combined with situations detected by other cameras, as described in the next section.

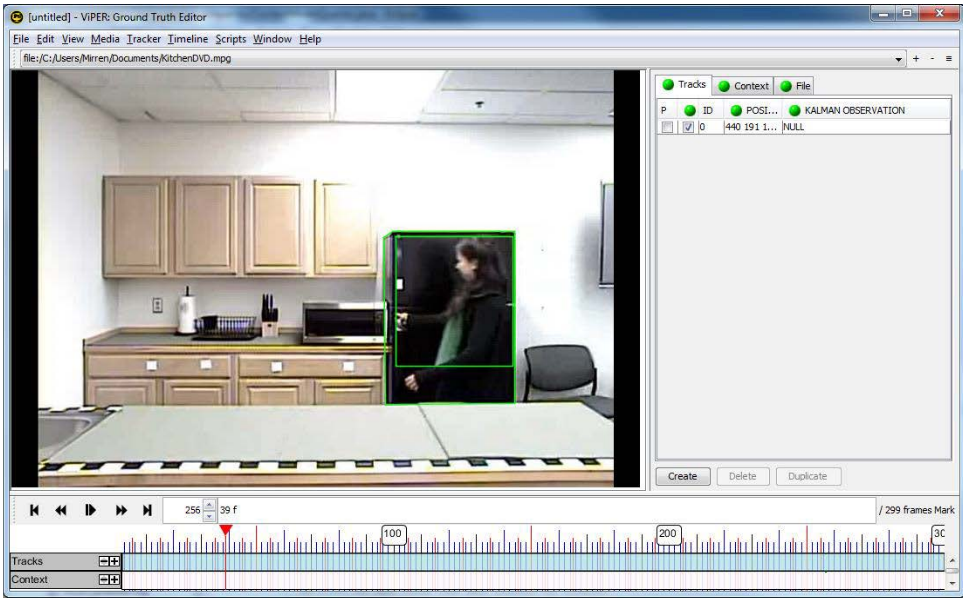


Figure 3.15: Camera 2: simple scene recognition (enclosing)

Similar rules have been defined for other cameras. For example, the rule defined for camera 1 (Fig. 3.16). In this case, the goal is detecting the overlapping between the person and the fridge bounding boxes, which is represented with the RCC predicate `PO` (partially overlap). In this manner, the system detects when the person is inside the fridge area, which frequently means that he or she is interacting with the object (Fig. 3.17).

```
;;; Single-camera simple scene recognition (camera 1) [4]
(firerule
       (and
         (?track        #!tren:Track)
         (?person       #!smarthome:Person)
         (?person     ?track          #!scob:hasAssociatedTrack)
         (and         (?object        #!smarthome:Fridge)
                      (?*object       ?*track     :po)))
  (
       (instance
         (new-ind ind ?person ?object)      #!smarthome:Touch)
         (related (individual ind) ?person #!smarthome:touch)
         (related (individual ind) ?object #!smarthome:touch))
   )
)
```

Figure 3.16: Simple scene recognition touch rule

3.8 Multi-camera scene identification

In the simple camera recognition example, it is described how a single camera detects the situation when a person is operating the fridge as a result of the instantiation of the RCC

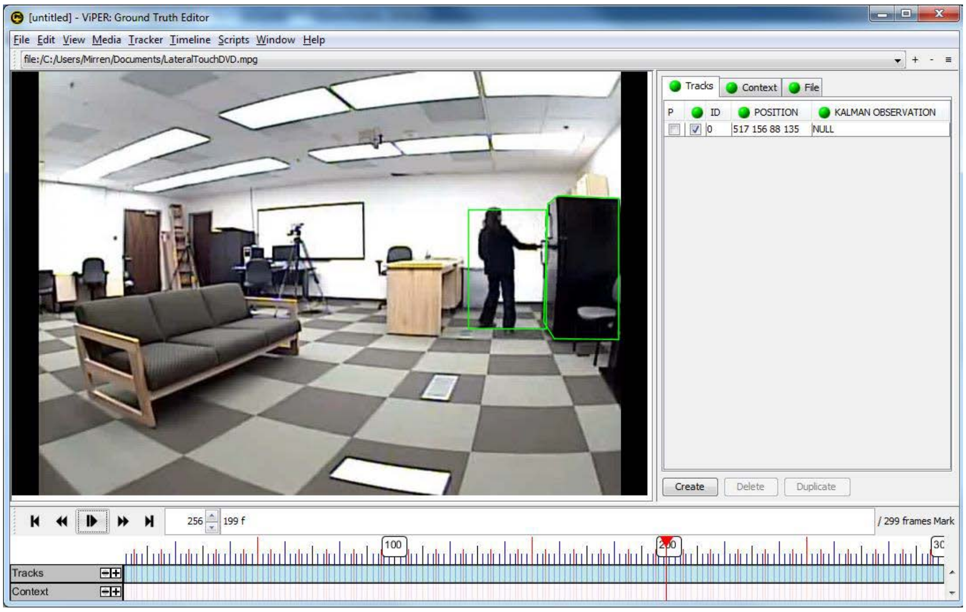


Figure 3.17: Camera 1: simple scene recognition (touch)

property `PO`. Nevertheless, this situation might be also detected when the person is in front of the fridge, because the rule antecedent is also true. As shown in Fig. 3.18, that results in the misinterpretation of a situation.

There are two main solutions for this problem. On the one hand, it is possible to perform a low-level calibration of the cameras and use a numerical procedure to fuse object positions in local coordinates acquired by different cameras to obtain a combined position in global coordinates. On the other hand, consistently with our architecture, it is possible to process local scene interpretations at the central node. The ACTV ontology is used to communicate local scenes to the central node. This information is encoded as instances of the `Situation` concept of ACTV, besides additional instances that may be interesting –e.g., the objects involved in the action. The `Situation` instances are tagged to identify the camera that has detected them with the `capturedBy` property. When the detected situations are received by the central node, they are also asserted as instances as the `Recent` concept, which includes all the situations in the current temporal window. Periodically, the central node runs an update procedure to retract situations as `Recent` and assert them as `NotRecent`, in such a way that they are marked as outdated and will be no longer able to trigger certain reasoning procedures.

After receiving situation information, the central node applies rule-based reasoning to discard or confirm the information provided by single cameras. In the example depicted in Figs. 3.15 and 3.17, the central node receives the situation information obtained by camera 1 and camera 2 at the same temporal window. Camera 1 informs of a `Touch` situation involving `person1` and `fridge1`. Camera 2 informs of an `Enclosing` situation involving `person1` and `fridge1`. A rule to create a proper `ConfirmScene` instance of the RECO ontology has been created in the context model of the central node (rule 3.19).

Notice that the rule implicitly assumes that there is only one person on the scenario. The

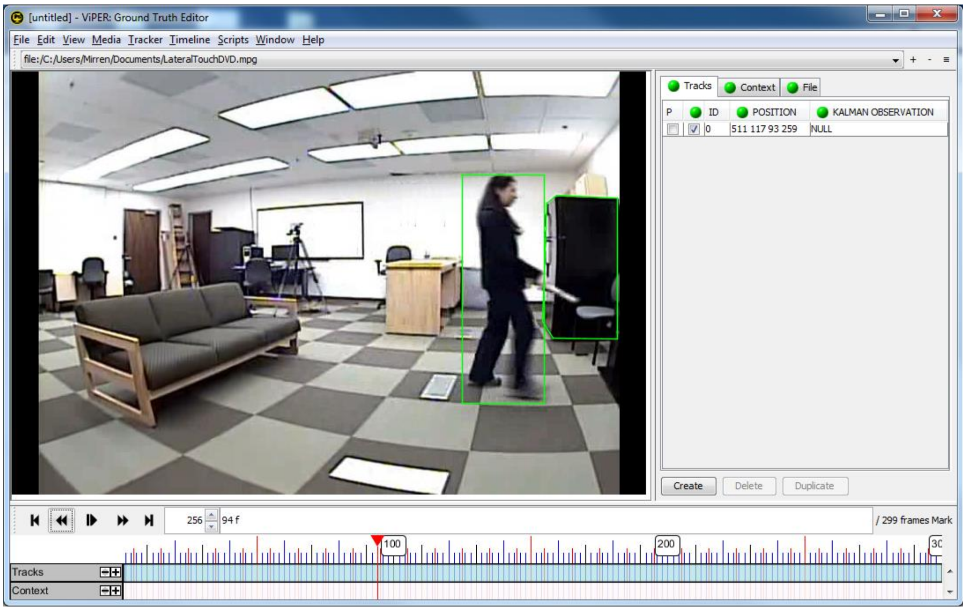


Figure 3.18: Camera 1: bad scene recognition (touch)

```
;;; Scene confirmation (central node) [5]
(firerule
       (and
         (?s1    #!smarthome:Touch)
         (?s1    #!tren:camera1           #!tren:capturedBy)
         (?s1    #!smarthome:fridge1      #!smarthome:touch)
  (?s2    #!smarthome:Enclosing)
         (?s2    #!tren:camera2           #!tren:capturedBy)
         (?s2    #!smarthome:fridge1      #!smarthome:enclosing)
  (?s1    #!reco:Recent)
  (?s2    #!reco:Recent)))
  (
       (instance
         (new-ind ind ?s1 ?s2)             #!reco:ConfirmScene)
         (related (individual ind) ?s1     #!reco:confirm)
         (related (individual ind) ?s2     #!reco:confirm))
   )
)
```

Figure 3.19: Central node sceneconfirmation rule

rule creates a new instance of `ConfirmScene` related with the situations sent by camera 1 and camera 2. In this case, no further processing is performed, since cameras are by default confident with their local interpretations. The fusion node behaves as a high-level tracker, since it calculates a better estimation of the position of the person from situation information provided by single cameras. The consequent of the rule can be easily extended to assert the confirmed scene as an instance of the `UsingFridge` class as well, thus creating a unified view of the scenario.

Likewise, similar rules can be created to discard scenes when they change to the `NotRecent` state and have not been confirmed. In this case, the cameras are notified to retract the unconfirmed situations from their context model. A `RetractScene` instance would be sent back to the cameras to adapt their behavior to the global situation, in a similar way as it has done for tracking enhancement. That means that the camera is recommended to remove the instances of its cognitive model representing the unconfirmed situation –for example, the `Touch` situation involving `person1` and `fridge1` in camera 1 detected in Fig. 3.18, thus preventing the execution of the rules with matching situation conditions in their antecedent. The local instantiation of the context model is therefore adapted without modifying the rule base.

These features of the central node are envisioned to provide support for more complex scene recognition procedures. For instance, imagine that camera 2 detects that a milk bottle has been left on the table after the `UsingFridge` situation. At this point of the execution, two previous situations are states in the knowledge model: a) the person was using the fridge (previous situation confirmed by camera 1 and camera 2); b) the bottle is on the table. A simple rule could be asserted: if the "person has been using the fridge" and the "bottle on the table", then it can new inferred that the person is preparing breakfast. Obviously, this rule is too simple and should be improved to avoid false positives (e.g., daytime can be also considered), but it shows the potential of the cognitive model and how the system can be compositionally extended with new situation detection heuristics. Extending and testing the framework to deal with hese situations is the most promising direction for future research.

4

Framework extensions for visual sensor networks: A Social Signal Processing example

Recent advances in technologies for capturing video data have opened a vast amount of new application areas in visual sensor networks. Among them, the incorporation of light wave cameras on Ambient Intelligence (AmI) environments provides more accurate tracking capabilities for activity recognition. Although the performance of tracking algorithms has quickly improved, symbolic models used to represent the resulting knowledge have not yet been fully adapted to smart environments. This lack of representation does not allow to take advantage of the semantic quality of the information provided by new sensors. This chapter advocates for the introduction of a part-based representational level in cognitive-based systems in order to accurately represent the novel sensors' knowledge. The chapter also reviews the theoretical and practical issues in part-whole relationships proposing a specific taxonomy for computer vision approaches. General part-based patterns for human body and transitive part-based representation and inference are incorporated to a the ontology-based framework previously presented in 2 and 3 to enhance scene interpretation in the area of video-based AmI. The advantages and new features of the framework are demonstrated in a Social Signal Processing (SSP) application for the elaboration of live market researches.

4.1 Introduction

AmI develops computational systems that apply Artificial Intelligence techniques to process information acquired from sensors embedded in the ambience in order to provide helpful services to users in daily activities. AmI objectives are: (i) to *recognize* the presence of individuals in the sensed scene; (ii) to *understand* their actions and estimate their intentions; (iii) to *act* in consequence.

The use of visual sensors in AmI applications has received little attention (Remagnino and Foresti, 2005), even though they can obtain a large amount of interesting data. However in the last decade, new visual sensor technologies have updated the established concepts of the computer vision approaches. Time-of-Flight (ToF) technology provides both intensity and distance information for each pixel of the image, thus offering 3-dimensional imaging (Foix et al., 2011), (Kolb et al., 2009). Structured light imaging allows to obtain an accurate depth surface for objects with an unprecedented resolution. Recently, the cost of these sensors has been dramatically reduced, which has lead to a widespread adoption of these technologies, now even present in consumer electronics like the Kinect^{TM} peripheral for Microsoft XBox^{TM} system.

Novel computer vision algorithms have been proposed to detect and track human movements from structured light and ToF sensors (Ganapathi et al., 2010). These works are mostly based on the definition of a model and motion of the human body. To name some application areas, ToF-based systems have been used in tracking algorithms for the detection of moving people (Kahlmann et al., 2007), nose detection algorithms (Haker et al., 2007), body gesture recognition (Holte et al., 2008), hand tracking proposals (Breuer et al., 2007), (Soutschek et al., 2008), SSP to classify human postures (Wientapper et al., 2009) and Ambient Assisted Living (AAL) to detect people falls (Leone et al., 2011).

Unfortunately, current approaches do not provide a well-defined model to represent the semantic details of the data, such as relationships or constraints, coming from new algorithms. The use of a conceptual model offers several advantages at a low cost. Formal models establish a common symbolic vocabulary to describe and communicate scene data while providing support for logic-based reasoning. Symbolic language is closer to human language, and therefore it is easy to interact and interpret system inputs and outputs. Reasoning, in turn, can be applied to check the consistency of the models and to infer additional knowledge from explicit information.

The formulation of models based on abstraction levels has led to the implementation of non-cohesive systems which are not able to fluently communicate among themselves. For this reason, it is necessary to provide new common and transverse knowledge layers among these levels including new semantic relationships. The goal of this stategy is the close interaction among semantically similar layers to the automatic generation of new knowledge. With the advent of new sensors, we advocate for the addition of a representation layer based on mereology and meronymy. Meronymy studies part-whole relations from a linguistics and cognitive science perspective. Mereology is a close concept, which concerns the formal ontological investigation of the part-whole relation and it is formally expressed in terms of first-order logic. The idea of employing a part-based layer to support the statements of the scene object abstraction level in a cognitive architecture has been previously suggested by Pinz et al. (Pinz et al., 2008). This proposal goes further and seeks to provide a symbolic layer based on the formal definition, development patterns and implementation of part-whole relationships.

Symbolic data representations allow to develop cognitive models able to represent more accurately the complexity of the scene. These models can analyze systematically the knowledge of the scene to discover and describe data related with activities developed by a subject fusing its representation with high-level context knowledge (see A). A key part of such analysis is currently supported by the approaches emerged from a cognitive view of the traditional computer vision techniques. The ties between meronymy and the current spatial qualitative approaches (Randell et al., 1992), (Allen, 1983) in cognitive vision –mainly focused on a qualitative description of spatio-temporal aspects (Renz, 2002)– must be regarded as crucial to narrow the gap of knowledge in activity recognition approaches.

This chapter describes an ontology-based model for data acquired from recognition algorithms through light wave technology. This model is incorporated into a cognitivist framework for contextual fusion of 2D visual information previously proposed in 2.

A general taxonomy of part-whole relationships for computer vision is proposed. The relationships are distributed along the levels of the model according to their abstraction. Several general pattern based on transitive part-whole relationships are proposed to cover the representation of the data to the level of accuracy currently achieved and to improve the quality of the inference process.

To illustrate the functioning of the extended framework a case study based on a SSP environment is presented. SSP aims at providing computers with the ability to sense and understand human social signals (Vinciarelli et al., 2009). The example depicts a novel application of structured light cameras for live market researches. The goal is the formal representation of complex activity recognition and the automatic reasoning through ontologies. The example incrementally describes the activities representation through the presented model and the automatic structuring of event knowledge along the part-based level. Straightforward rules corresponding to a logic inference engine are attached to the example sections to demonstrate the feasibility of the proposal.

4.2 Theoretical issues in part-based representations

Meronymy has been subject of researches in linguistics, philosophy, and psychology. From a philosophical point of view parts have been characterized as single, universal and transitive relations used to model, among others, the spatio-temporal domain (Girju et al., 2006). This definition stay open since it was criticized using an axiomatic representation which considers part-of a partial ordering relation (Simons, 1987). Afterwards the representation was completed with the addition of new axioms (Simons, 1991).

Representations of part-based relations are founded on the Ground Mereology theory. The Ground Mereology establishes three principles (Varzi, 2011):

- Reflexive: Everything is part of itself.
  $\forall x(part\_of(x,x))$
- Antisymmetric: Two distinct things cannot be part of each other.
  $\forall x,y((part\_of(x,y) \wedge part\_of(y,x)) \longrightarrow x = y)$
- Transitive: Any part of any part of a thing is itself part of that thing.
  $\forall x,y,z((part\_of(x,y) \wedge part\_of(y,z)) \longrightarrow part\_of(x,z))$

These principles have been a source of discussions in meronymy due to the need to consider different kinds of part-whole relations and because of some of them must be intransitive. Some examples can be found in (Odell, 1994).

The variety of semantic senses in part-whole relations drove researchers to look for a collection of part-whole relations. Winston et al. (Winston et al., 1987) developed a taxonomy founded on three linguistic and logical characteristics: functional, homeomerous and separable. These characteristics define a set of six meronymic relations: component-integral object, member-collection, portion-mass, stuff-object, feature-activity and place-area.

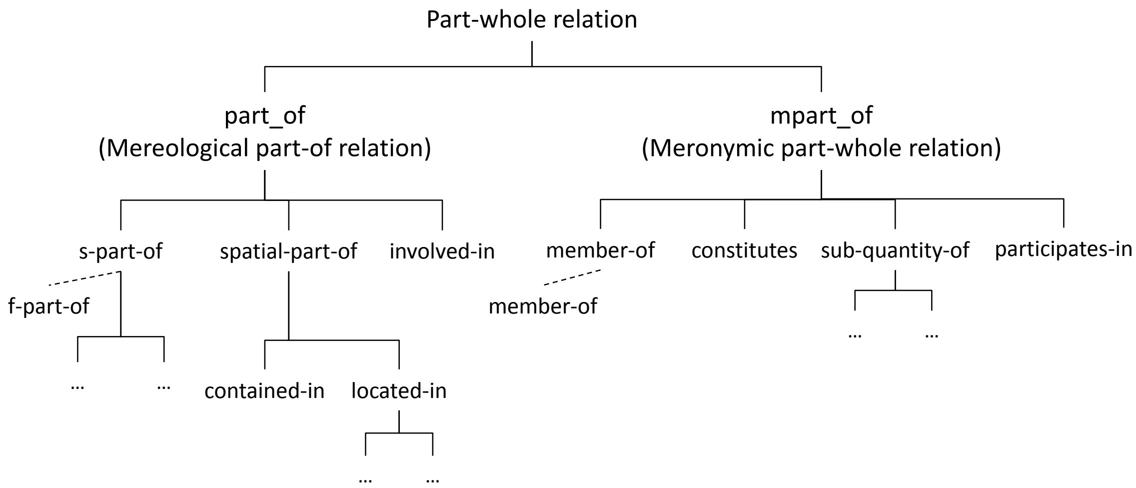


Figure 4.1: Keet's et al. taxonomy of basic mereological and meronymic part-of relations

Keet et al. (Keet and Artale, 2008) proposed a formal taxonomy of part-whole relations which implements a compromise solution for the "ontologically-motivated relations useful for conceptual modeling up to the minimum level of distinctions". This taxonomy is particularly relevant since the properties are defined using categories of the DOLCE (Gangemi et al., 2002) upper ontology. The taxonomy by Keet et al. is extended in Section 4.4.1 to be applied in cognitive vision environments.

Interestingly enough, connectedness is a fundamental concept shared between the foundations of mereological and topological theories. As it is shown in mereotopological approaches (Varzi, 1996) topology can be defined as a domain specific subtheory of mereology and mereology can be defined as a subtheory being topology primal. An example of the latter is the theory developed by Randell et al. (Randell et al., 1992), who propose the Region Connection Calculus (RCC) (see B.1).

Current capabilities in computer vision systems do not allow an easy recognition of mereological relationships from spatial inclusion assertions. Topological relationships between two entities, for example, TPP, NTPP, EQ or PO relations are essential clues to detect part-whole patters; however, it is also necessary the detection of a connection relation among the content and the container. On the other hand, this proposal advocates for the combined used of spatial and mereological knowledge at different levels. A separate definition of theories can be used to classify and assert new knowledge. A clear example is the classification of subactivities. The spatial context of a subactivity can determine the relationship with the overall activity. Comparing products into the supermarket is part of shopping; however, comparing products can be part of cooking if the subject is into a kitchen. Sections 4.5.1 and 4.6.1 present a practical approach on the combination of topological and mereological relations and their implementation in our system.

4.3 Ontology-based computer vision model and light wave technology integration

Changes are needed to model tracking data coming from novel devices. The priority to adapt these changes is to maintain the compatibility with the previous approaches. The ontologies of the initial framework have been extended to include support for light wave data:

- An additional Euclidean dimension for the depth position of recognized objects. This is easily achieved by relying on the qualia approach (Gangemi et al., 2002) used in the original ontology model to represent properties and property values.

- A new definition of the concepts that represent human entities in the scene. Essentially, the current description of a subject in the scene, represented by the `Person` concept, is now associated with a description of anatomical joints and limbs. This description has been formalized according to existing patterns to represent part-whole relations with ontologies and current ToF-based computer vision models for articulated bodies.

The introduction of new devices requires upgrading the capacity of spatial representation in the model from two to three dimensions. These changes concern both perceptual data captured by light wave cameras and context data representing physical objects. The previous model followed the qualia approach used in the upper ontology DOLCE (Gangemi et al., 2002). This modeling pattern distinguishes between properties themselves and the space in which they take values. The values of a quality –e.g. `Position`– are defined within a certain conceptual space –e.g. `2DPoint`. To adapt the ontology-based model to this new quality space, the `3DPoint` concept, which represents a position using three coordinates, is included as a subclass of `PositionValueSpace`, which represents the space of values of the physical positions.

Current Kinect™ algorithms are able to detect real-world entities; e.g. a person including data related to the human limbs and joints. Our ontology-based model represents these kind of real-world data at the scene object level. However, these data also include low level information that should be represented as tracking entities to support the scene object assertions. Tracking entities level has been adapted to represent low level data of human members and joints –position, size, kinematic state, and so on– this information is associated to the `Person` concept declared in the scene object level. The inclusion of limbs and joints is compliant to the previous version of the tracking entities ontology. The applied part-whole pattern (see Section 4.4.2) allows keeping backward compatibility. In fact, this model can combine 2D monocular cameras and light wave devices using the same set of ontologies.

4.4 Part-based symbolic layer for cognitive vision approaches

This section presents a part-based taxonomy of properties for cognitive vision environments based on some approaches discussed in Section 4.2. Afterwards a general ontology-based pattern to represent the transitive properties of the taxonomy is explained. The semantics of the human body and its parts are used illustrate this pattern. Thereby a dual purpose is fulfilled: the explanation of the general pattern and its application to exploit the detection of human body structures using novel devices.

4.4.1 Part-based taxonomy of properties for cognitive vision environments

The identification of the underlying characteristics presented in Section 4.2 allows to discriminate between several kinds of part properties. The characteristics by Winston et al. are appropriate for cognitive vision representation because they are mainly supported by spatio-temporal foundations. However this set of characteristics is too small and do not allow a wide specialization of properties. Thus we have also taken into account the classification by Opdahl et al. (Opdahl et al., 2001) (see Table 4.1).

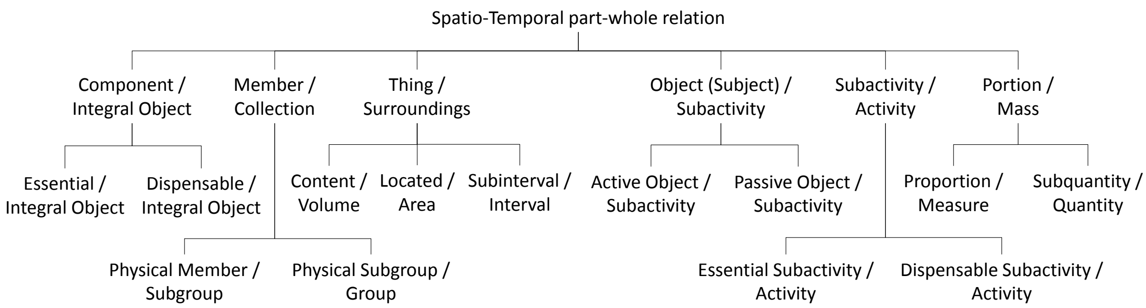


Figure 4.2: Proposed taxonomy of part properties for spatio-temporal aspects in vision-based systems

The resulting classification is focused on properties which can be projected as spatial and temporal concepts captured by visual devices. Fig. 4.2 shows the proposed taxonomy taking into account the spatio-temporal aspects in vision-based systems. Below we carry out an analysis based on characteristics of part properties. This analysis only considers the general characteristics of each property. An exhaustive list of characteristics is not offered for each property because some of them do not characterize the property. Current classification can be reconsidered for a specific specialization according to a particular domain. It is considered that all the properties meet the Ground Mereology principles except transitivity.

Component/Integral object (`componentOf`): This is a functional, separable, resultant and transitive property. The property is relevant for unidentified entities and scene objects. Thus it is mandatory to define a set of subactivities where the part can intervene. There are two subtypes: (i) Essential/Integral object (`essentialComponentOf`) are those critical parts to identify a whole, for example, the chest of a body. Their characteristics, in addition to the inherited, are: mandatory, existential dependency and immutabe. (ii) Dispensable/Integral object (`dispensableComponentOf`) are those parts that are not crucial for recognition. Following the previous example, a hand can be regarded as a dispensable component for body recognition. Their corresponding characteristics are: optional and mutable.

Member/Collection (`memberOf`): This property aims to redefine the identity of an entity through its assimilation to a group. The necessary characteristics of this property are separable, optional, mutability, shareability. Generally this property is intransitive when it is used for abstract sets of membership, for example, when a person is part of an organization.

Table 4.1: Set of characteristics to classify part-whole relations

Characteristic	Definition
Functional	Parts are in a specific spatial/temporal position with respect to each other supporting their functional role with respect to the whole.
Homeomerous	Parts are visually similar to each other and to the whole to which they belong. Parts and aggregates belong to the same class.
Separable	Parts can be physically disconnected from the whole to which they are connected and can be detected without being part of a particular aggregate object. The opposite characteristic is **Invariance**.
Resultant	A part provides at least one property that extends to the whole.
Mandatory	An object of a particular class must be detected to declare the existence of an aggregate object. The opposite characteristic is **Optional**.
Existential dependency	A single and always the same occurrence of an object is critical for the life of the aggregate.
Mutability	A particular part object can be replaced in the aggregate object by another equivalent part without losing its identity. The opposite characteristic is **Immutability**.
Shareability	An object can be part of more than one aggregate object at the same time.
Transitivity	An object A is part of an aggregate B, the aggregate B is in turn part of another aggregate C, then A is also part of C. The opposite characteristic is **Intransitivity**.

The subproperties are specialized in the spatio-temporal level where can be detected according to proximity measures or similar kinematic features: (i) Physical member/Subgroup (`physicalMemberOf`) which meets the mandatory characteristic because the parts only can be scene objects corresponding to context data or detected entities with physical features; (ii) Physical Subgroup/Group (`physicalSubGroupOf`) which meets transitivity, homeomerousity and mandatory characteristics because parts only can be clusters of physical members.

Thing/Surroundings (`settledIn`): This property defines a content relationship and an invariant connection between the part and the whole. It is only applicable between objects and entities with spatial or temporal representation. The general characteristics of this property are: homeomerousity, invariance, optional, immutability, shareability and intransitivity. The transitive, mandatory and existencial dependency subproperties are: (i) Content/Volume (`containedIn`) is exclusively used by spatial representations based on 3D points; (ii) Place/Area (`locatedIn`) is exclusively used by spatial representations based on 2D points; (iii) Subinterval/Interval (`intervalOf`) is used by temporal representations based on time intervals.

Object (Subject)/Subactivity (`involvedIn`): This intransitive property defines the subjects who are involved in an activity. Its characteristics are functional, non homeomerous, separable, optional and shaerable. Objects and subjects with functional part properties in their definition are the main candidates to instantiate this property. The identified subproperties are not based on any characteristic but in our knowledge about the activity recognition: (i) Active Object/Subactivity (`activelyInvolvedIn`) is instantiated when the object performs the activity; (ii) Passive Object/Subactivity (`passivelyInvolvedIn`) is instantiated when the object is passively involved in the activity.

Subactivity/Activity (`participatesIn`): Represents the relation among straightforward activities which participates in more complex activities. The main characteristic of this property are: functional, separable, homeomerous, transitive and shaerable. The property can be divided in: (i) Essential Subactivity/Activity (`essentialSubActivityOf`) if the subactivity is mandatory for the recognition of a more complex activity. Its specific characteristics are: mandatory, existential dependency and immutability. (ii) Dispensable Subactivity/Activity (`dispensableSubActivityOf`) if the subactivity is not crucial to recognize a more complex activity. Its specific characteristics are: optional and mutability.

Portion/Mass (`portionOf`): Necessary characteristics of this property are: homeomerousity, separability and intransitivity. Two transitive subproperties have been identified: (i) Proportion/Measure (`proportionOf`) if the property is countable with a spatio-temporal measure. For example, a second is the sixtieth part of a minute. The corresponding characteristics are: functional, mandatory and existential dependency. (ii) Subquantity/Quantity (`quantityOf`) if does not exist a visual proportion between the part and the whole. For instance, the part of the water spilled from a cup. The inherent characteristic of this subtype is mandatory.

Stuff/Object (`madeOf`): The constituent material can help to identify an object avoiding false positives in the entity detection process. This property is typically used in part-based taxonomies; however it can not be detected in the scope of vision systems.

Some other characteristics from (Opdahl et al., 2001) classification have not been mentioned because they are already defined in the (Winston et al., 1987) set of properties (e.g. abstraction and homeomerousity), have the same name but a different meaning (e.g. separability) or are not general (e.g. shareability). It is interesting to note that shareability can be seen as a cardinality restriction for specific cases of some relationships. For example, a chest only can be part of one body. These kind of situations become a problem if the relationship is transitive. In Section 4.4.2 we present a pattern to manage the semantic of these situations.

Some of the properties shown in the previous taxonomy are intransitive, for example, `involvedIn` and `physicalMemberOf`. Sometimes there are complementary transitive relations that can be used to propagate a property along another property. The corresponding properties of the previous examples would be `participatesIn` and `physicalSubGroupOf`. To illustrate this, let us suppose a person who is a physical member of a group and the same group is part of a bigger group. This procedure only requires to declare the `physicalMemberOf` property along the `physicalSubGroupOf` property to automatically assert that a person is a physical member of the bigger group. A wider and strongly related vision of this issue is the table developed in (Sattler, 1995) which defines the conditions for the overall set of transitive interactions between different types of properties.

4.4.2 General model for ontology-based human skeleton representation

There are several existing ontologies designed to share and reason with structured data representing human anatomy (Rosse and Jr, 2003). Unfortunately, these ontologies have been developed in biomedical environments and define a complex conceptualization which is not useful to our needs. There are also other ontologies that represent the human body in a more simplified way (Gutiérrez et al., 2007); however these ontologies are not designed to deal with sensor data in a cognitive environment. A general pattern based on part-whole relationships is proposed to cover the semantic representation of data captured using light wave sensors. The designed ontology adapts the patterns presented in (Rector et al., 2005) and follows the conceptualization of articulated bodies shown in (Knoop et al., 2005) while keeping compatibility with DOLCE. The proposal can be broadly adapted to other fields.

Real-world knowledge achieve a more comprehensive representation organized through mereological relationships. A clear example is how the human mind divides the structure of a body in subjective parts. The current capabilities of Kinect™ skeletal view (see Fig. 4.3[1]) allow the description of a detected person in terms of two kinds of attributes: (i) body members –hands, feet, thigh, and so forth; (ii) joints –shoulders, elbows, wrists, knees, and so forth. A conceptualization of the attributes detected and the limbs composed by these attributes is represented in the tracking entities level. Resulting concepts represent the parts of the human body which are embodied in the `Person` concept.

The properties named below (`partOf` and `partOf_directly`) correspond to the `componentOf` subtype of properties. The names have been modified to present the pattern in a general way since it can be applied to the rest of properties defined in Section 4.4.1.

Two properties are used to represent the part-whole relationships: (i) `partOf`; (ii) `partOf_directly` –a `partOf` subproperty. `partOf` is a transitive property whose goal is establishing the correspondences between the parts and all the entities containing them.

[1] Fig. 4.3 source. http://embodied.waag.org

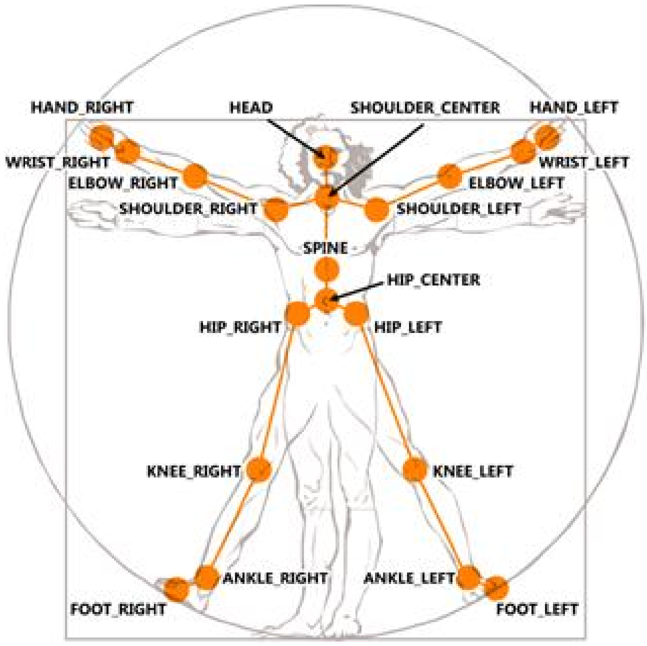


Figure 4.3: Joints captured by Kinect™ skeletal view

partOf_directly defines the subjective relation among a part and the next direct level of composed entities. These properties are necessary since cardinality restrictions over transitive properties, such as partOf, are not allowed by OWL-DL. Therefore, partOf_directly is used to define restrictions to maintain cardinality consistency, and partOf is used to infer both direct and indirect parts by means of transitivity and partOf_directly property instances.

The previous ontology is extended with classes to represent direct parts –e.g. PersonPartDirectly– and the overall set of part-whole relationships –e.g. PersonPart. PersonPartDirectly subsumes direct parts of a Person such as Head, UpperLimb and LowerLimb. The classes hosting direct parts state existential range restrictions over partOf_directly properties –e.g. partOf_directly some Person. On the other hand PersonPart subsumes the set of parts of the Person concept. In this case, the direct parts of an UpperLimb concept, namely Arm, Forearm, Hand, Shoulder, Elbow and Wrist, are classified as subclasses of PersonPart; however they are not considered subclasses of PersonPartDirectly. The classes hosting direct and non direct parts state existential range restrictions over partOf properties –e.g. partOf some Person.

To improve the consistency, cardinality restrictions –exactly 1– are stated over partOf_directly as necessary conditions into the concepts corresponding to body members and joints. This means "a part only belongs directly to the next level entity and just to that entity".

The combined use of the part properties and the restricted classes leads reasoners to automatically infer new taxonomies derived based on part-whole relationships. Fig. 4.4 illustrate an example of a taxonomy inferred from a explicitly stated taxonomy. Unfortunately, adding cardinality restrictions on each concept could significantly affect the performance of the reasoner. Some other configurations for this pattern are possible and also valid. This implementation tries to reduce the classification time while complying to the semantics of the human body domain.

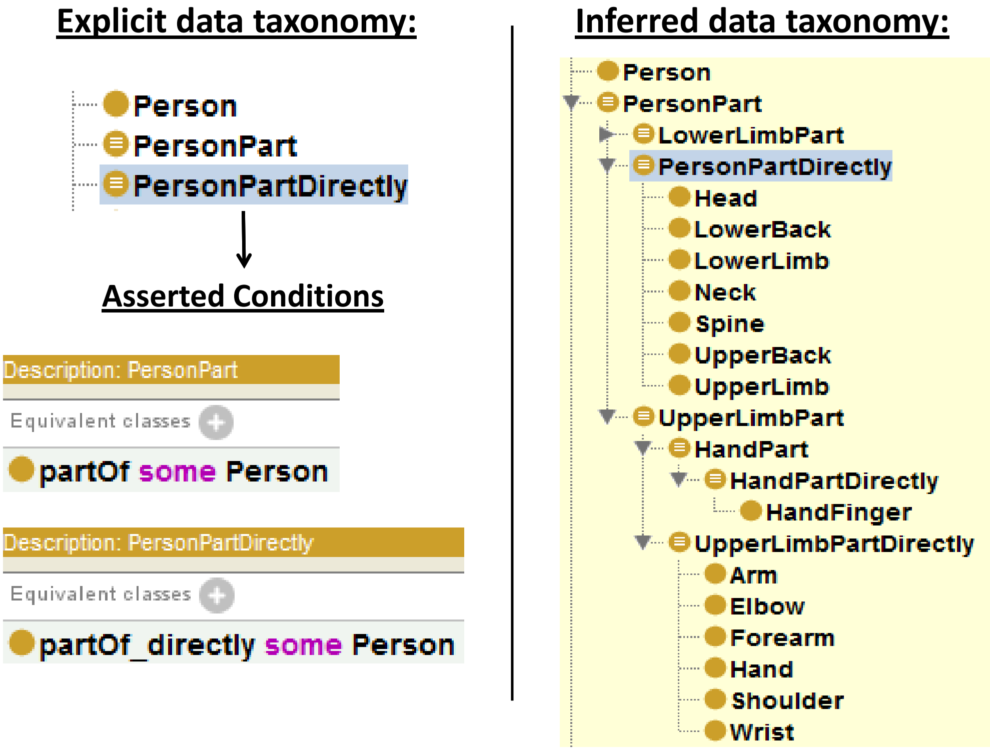


Figure 4.4: An example of explicit and inferred taxonomies

Considering the combination of the taxonomy presented in Section 4.4.1 and the pattern above, we obtain a taxonomy to tackle with the spatio-temporal issues of a cognitive vision system. Fig. 4.5 shows the implemented taxonomy, notice that some of the transitive properties do not include a direct property because it is implicit when the superproperty is transitive, for example, `dispensableComponentOf` and `essentialComponentOf` are regarded as direct properties because `componentOf` is transitive. Each subtaxonomy of properties is assigned to one or several level forming a transverse layer through the model showed at the beginning of Section 4.3.

The classification of joints is inspired by the virtual model shown in (Knoop et al., 2005). There are three types of joints (see Fig. 4.6) depending on the degrees of freedom (DoF): (i) `UniversalJoint`, three DoF; (ii) `HingeJoint`, one DoF and two restricted DoF; (iii) `EllipticJoint`, three restricted DoF. Joint concepts store important data such as the articulated body members and the angle between them. These data is basic to mantain the consistency and to improve the semantic capacity of the model.

The model is designed by taking into account future changes in the granularity of the obtained data. New devices able to offer an accurate definition of the body members –e.g the fingers of a hand– are easily adaptable. The larger the number of levels in the model, the greater amount of data is inferred.

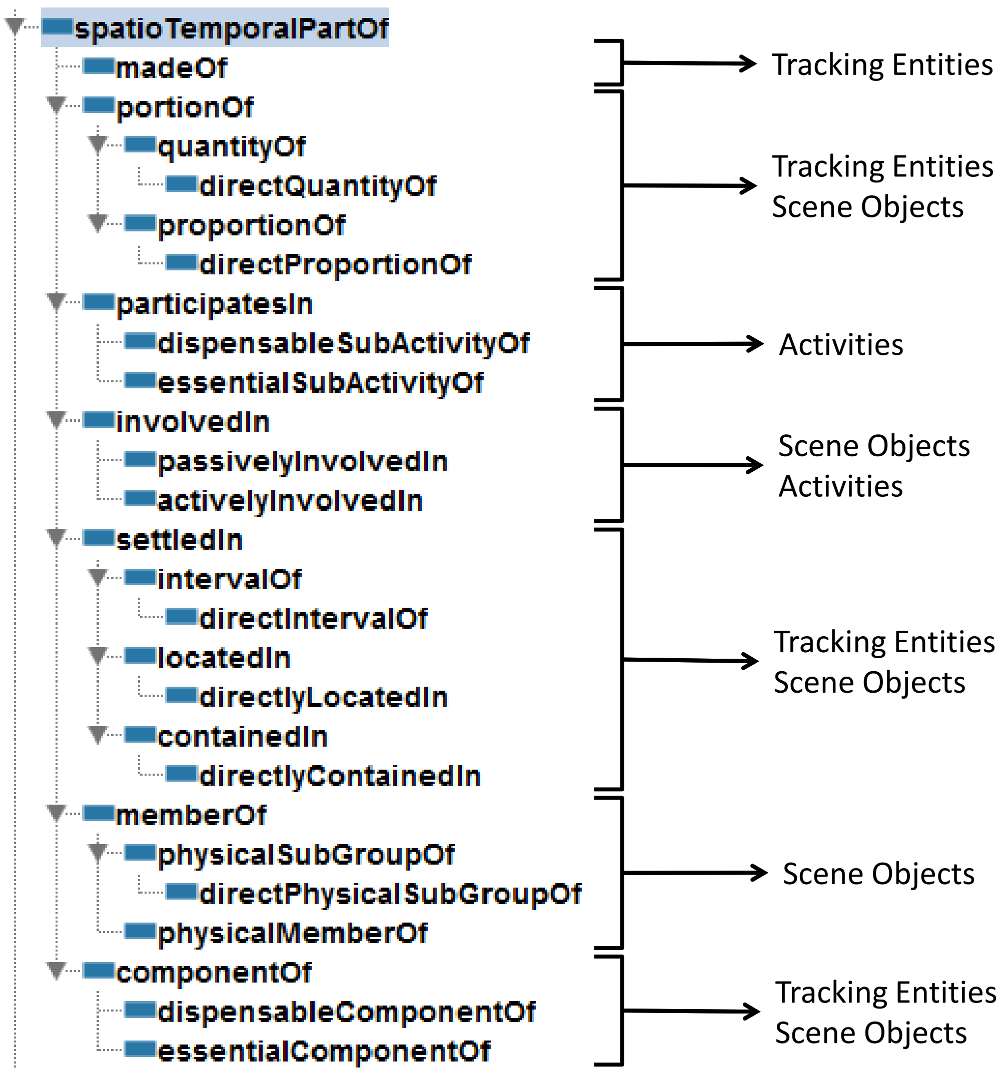


Figure 4.5: Spatio-temporal taxonomy with pattern representation

4.5 Part-based data extraction and propagation

There is an important amount of implicit knowledge surrounding the part-based approaches which should be extracted and used as a basis of the cognitivist models to improve the semantic richness and robustly justify the knowledge base reasoning.

4.5.1 Expliciting hidden relationships between subclasses, parts and locations

The research by Winston et al. (Winston et al., 1987) shows the power to find implicit relationships using deductive reasoning based on syllogisms. The conclusion of this study indicates that there is a hierarchical ordering respectively between class inclusion, mereological inclusion and spatial inclusion which implies that "syllogisms are valid if and only if the conclusion expresses the lowest relation appearing in the premises". Syllogism are a kind logical

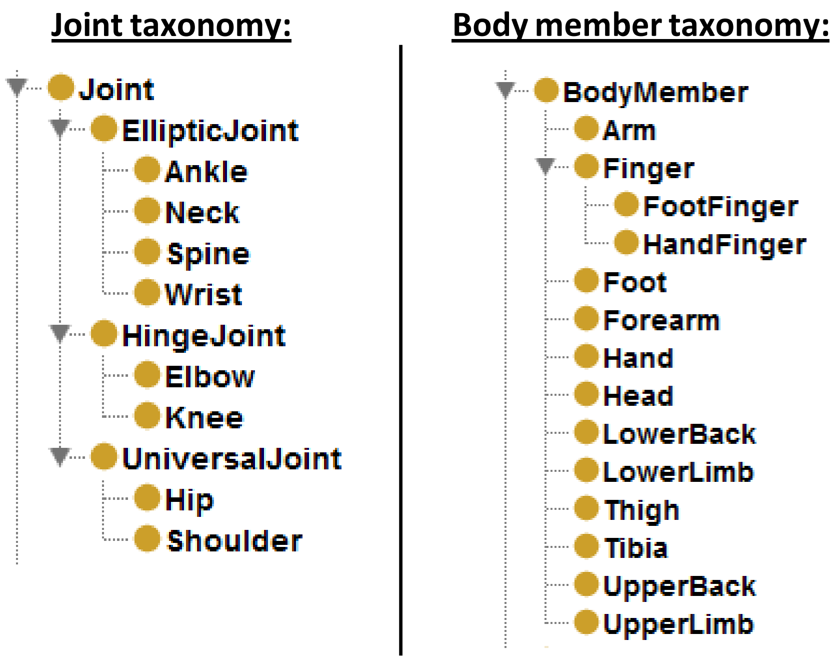


Figure 4.6: Explicit taxonomies for joints and body members

argument in which one proposition is inferred from two or more premises. A huge quantity of implicit relations can emerge from these inferences. The following example illustrates these assertions:

(1a) Peter is a physical member of a tourist group. (Mereological inclusion)

(1b) The tourist group is into the shop. (Spatial inclusion)

(1c) Peter is into the shop. (Spatial inclusion)

Ontologies have several advantages to carry out this kind of deductive reasoning because of: (i) the hierarchical structure of ontologies is strongly related to the idea of class inclusion since terminological boxes represent concepts as general classes which host more specific or specialized classes; (ii) the mereological patterns to represent and reason with parts and the current reasoner's support for qualitative spatial approaches (Stocker and Sirin, 2009) provide the semantic support to apply this kind of arguments; (iii) the OWL 2 construct `ObjectPropertyChain` allows a property to be defined as the composition of several properties. Compositions enable to propagate a property (e.g.; `placedIn`) along another property (e.g.; `partOf`). The previously described syllogism is automatically handled by the following statement:

```
SubPropertyOf( ObjectPropertyChain( :partOf :placedIn) :placedIn)
```
[2]

[2]Composition feature in OWL 2. http://www.w3.org/2007/OWL/wiki/New_Features_and_Rationale#F8:_Property_Chain_Inclusion Last accessed December 2015

Table 4.2: Composition of properties

⊗	**hasClass**	**partOf**	**placedIn**
hasClass	hasClass	partOf	placedIn
partOf	partOf	partOf	placedIn
placedIn	placedIn	placedIn	placedIn

Table 4.2 (Hornsby and Joshi, 2010) shows the syllogisms hierarchical ordering described through properties composition. Notice that the table's main diagonal compositions do not need to be declared since the properties are transitive.

4.5.2 Automatic data propagation of events of interest

Sometimes the knowledge originated in a entity component should be represented as knowledge directly attributable to the overall entity. A pattern for data propagation along the parts and to the whole can be deployed based on the pattern explained in 4.5 Another pattern from (Rector et al., 2005) is adapted to distribute the data concerning the events developed in the human body members. This pattern requires: (i) the creation of the `hasEvent` property, which indicates that a subject is the source of an event –these property can be also specialized to address more specific events; (ii) new classes –e.g. `EventInBody` or `EventInUpperLimb`– to classify events, which comprises all the events carried out by the body and their parts; (iii) the characterization of the `partOf` property as reflexive. As it is shown in 4.2 reflexivity is one of the principles of Ground Merology theory and dictates that "Everything is part of itself". These principles allows to include the whole entities in the taxonomy of parts. This causes the subsumption of the `Person` concept by the `PersonPart` class.

Classes which host instances of events state existential range restrictions over `hasEvent` properties, for example, `EventInBody` declares the restriction `hasEvent` *`someValuesFrom`*[3] `PersonPart` and `EventInUpperLimb` states `hasEvent` *`someValuesFrom`* `UpperLimbPart`. To illustrate this, let us suppose the detection of an event in a hand. After the instantiation of the event and the corresponding property `hasEvent`, the reasoner propagates the event to the `EventInBody` and `EventInUpperLimb` classes. Thereby, events are classified by following an organization refined by anatomical levels. In addition, this pattern represents the affirmation "an event carried out by a person is an event executed by the person or any of its parts".

This approach can be extended using a composition between the properties `componentOf` and `participatesIn`. Based on the relationship between an event and a body part, the relationships between parts of higher order that contains them and the event are automatically inferred. The following example syllogism and the Fig. 4.7 depicts this extension:

(2a) Upper limb is component of Robert. (Explicit)

[3] *`someValuesFrom`* restriction. `http://www.w3.org/TR/2004/REC-owl-features-20040210/#someValuesFrom` Last accessed December 2015

(2b) Robert's upper limb participates in embraces a lamp. (Explicit)

(2c) Robert participates in embraces a lamp. (Conclusion)

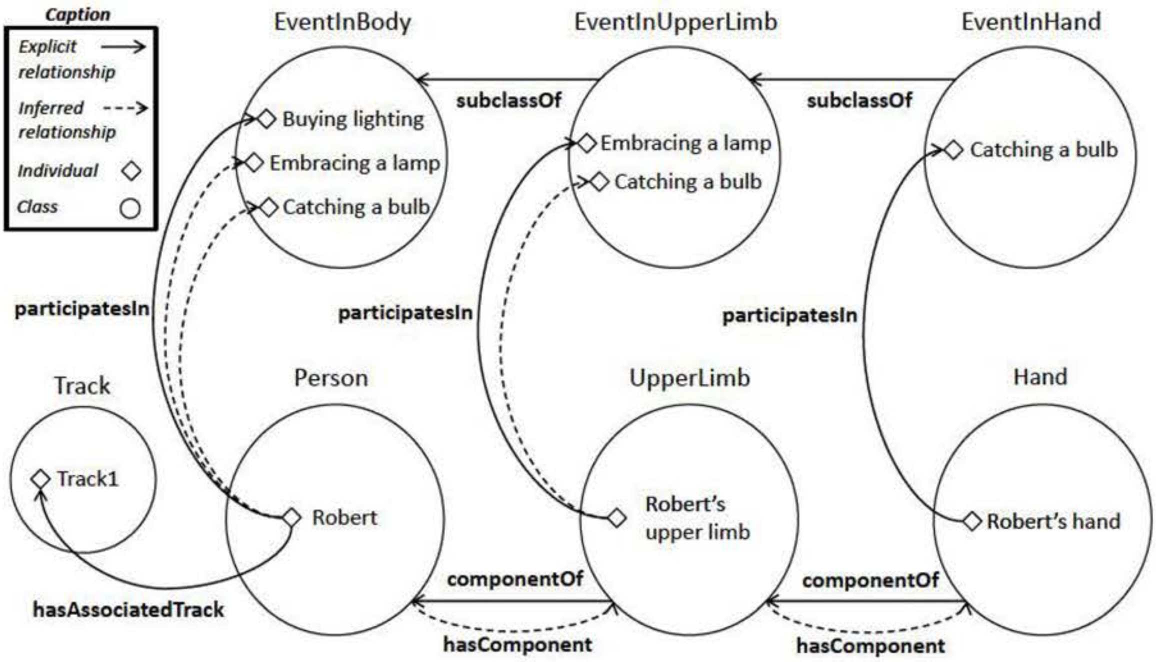


Figure 4.7: Inferred properties using composition between hasEvent and partOf

4.6 Implementation

The architecture presented in Section 4.3 has been implemented as a system prototype. The system's basic inputs are three: a variable amount of a priori knowledge, sensor data coming from different information sources and data formalisms represented with ontologies. The ontologies include a set of terminological boxes (TBoxes) each of which contains sentences describing concept hierarchies. In turn, an assertional box (ABox) contains facts about individuals of the domain of discourse. These TBoxes make up the structure of the vision-based AmI symbolic representation. The ABoxes of these levels are filled with assertions from predefined context knowledge, previous inferences and sensor data.

The overall system is based on the RACER[4] reasoner. The reasoner hosts the levels of the ontology-based computer vision model explained in 2.2; namely, tracking entities, scene object and activities. RACER has been chosen because it includes support for different kind of inference rules through the new Racer Query Language (nRQL), such as deductive, abductive, spatial and temporal (Gómez-Romero et al., 2011b).

Beyond the standard ontology reasoning mechanism based on subsumption, RACER also supports abductive and deductive rule-based inference. During the execution, abductive

[4]RACER engine web page. http://www.ifis.uni-luebeck.de/~moeller/racer/ Last accessed December 2015

nRQL rules defined in a subontology create new instances that are asserted into the same level or into an upper level. Eventually, the creation of new instances as defined in the consequents of the rules draws instances corresponding to an interpretation of the scene in terms of the activity ontology. Deductive rules, in turn, are used to maintain the logical consistency of the scene. The consistency verifies whether all concepts in the TBox admit at least one individual in the corresponding ABox.

The output of the system is a coherent and readable interpretation of the scene logically justified from the low-level data to the high-level interpretation.

4.6.1 Spatio-temporal support

RACER is the first inference engine able to manage the spatial knowledge through an implementation of the RCC (Randell et al., 1992) (see B.1 as an additional substrate layer. A substrate is a complementary representation layer associated to an ABox. The RCC substrate offers querying facilities, such as spatial queries and combined spatial and non-spatial queries. Although spatial instances from the ABox are not automatically connected with the RCC substrate, there is an identifying correspondence between them and the objects stored in the substrate.

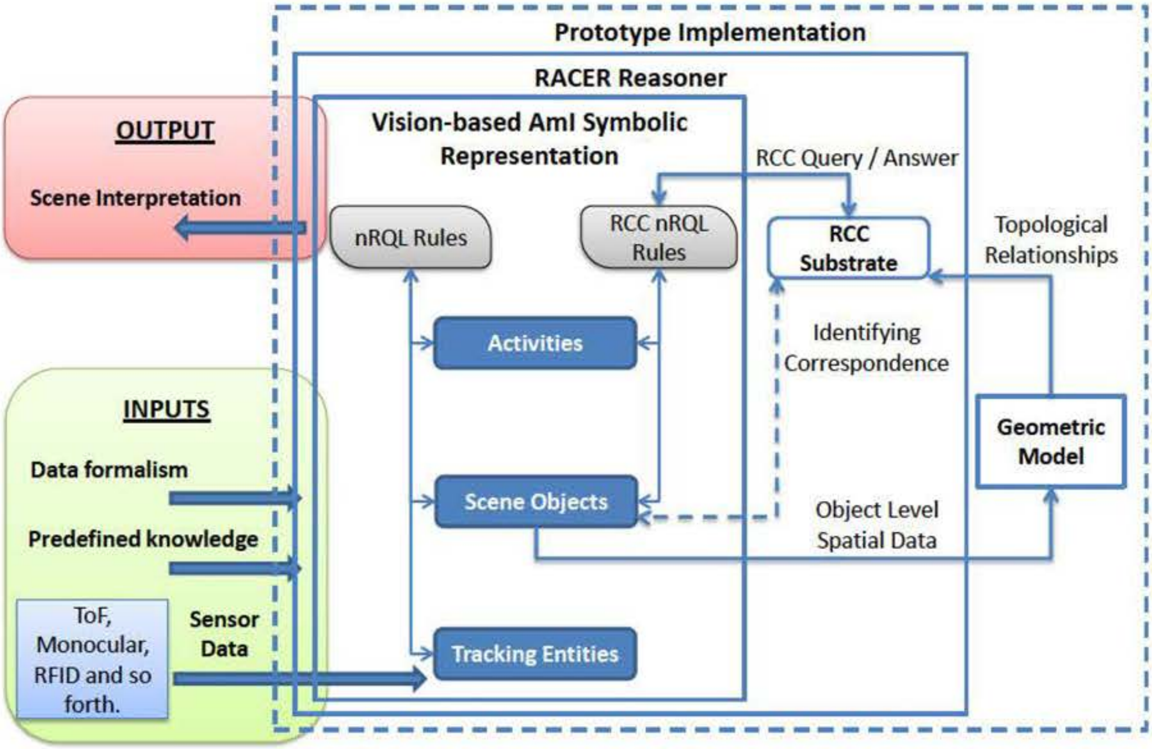


Figure 4.8: System implementation

A significant amount of knowledge of scene objects and activity levels is obtained by abductive rules that include spatial properties in their antecedent. Fig. 4.8 shows the integration of a geometric model in the system to dynamically calculate qualitative spatial relationships between scene objects. The geometric model receives spatial data from the scene object level. These data is instantiated into the Java Topology Suite (JTS)[5]. The JTS is an open source Java software library of two-dimensional spatial predicates and functions compliant to the Simple Features Specification SQL published by the Open GIS Consortium. JTS represents spatial objects in a Euclidean plane and obtains spatial relationships between two-dimensional objects quickly. Although OpenGIS spatial predicates and RCC-8 are not directly compatible, the output from the geometric model can be easily mapped from the OpenGIS format, in some cases, it only involves translating the name of the relationships. A correlation table between OpenGIS spatial predicates and RCC-8 can be found in (Schuele and Karaenke, 2010).

Additional improvements could be implemented to increase the computation speed. It is interesting to highlight that checking object spatial relations, and particularly RCC relations, has a complexity $O(n^2)$ –the test must be performed between each pair of elements. Thus, it would be convenient to build a data structure able to maintain a hierarchical spatial partition on the Euclidean space. Currently, our framework does not support these improvements, which remains as a promising line for future work (Serrano et al., 2011).

The temporal dimension is represented as timestamps and time intervals. As it is stated in 2.2.1 timestamps are represented using snapshots of capturing data. Time intervals representation is directly supported by the RCC substrate thanks to their proper relationships (Gómez-Romero et al., 2009). The temporal dimension can be applied in both ways into the antecedent of rules.

4.7 Case study: Live market research

Learning about relationships between the customer and the product at the point of sale is a very interesting knowledge in many economic fields, such as sales or marketing. Body gestures and spatial relationships contain useful knowledge about the sensations and intentions of shopping experiences. The model hereby presented can be used to automatically build live market researches based on the reactions and interactions of customers with the products.

Next subsections describe our system instantiation procedure and the expressiveness of the ontology model by presenting an activity recognition representation and a data propagation example. These subsections are depicted with rules to show its applicability in real environments.

4.7.1 Gesture instantiation procedure

A data set containing the skeleton representation of several –11– people was designed to test the new representation. These body structures were captured by using a Kinect™ sensor. For each person five types of upper limbs gestures were stored: down, open, up, diagonal and akimbo. A control system based on the OWL API[6] functionalities automates the assertion of data in the form of axioms from the capture device to the ontology formalism. The control

[5] Java Topology Suite web page. `http://www.vividsolutions.com/jts/JTSHome.htm` Last accessed December 2015

[6] OWL API web page. `http://owlapi.sourceforge.net/` Last accessed December 2015

system manages the classification of the individuals received from the Kinect™ sensor, the explicit property instantiations such as `partOf_directly` and the instantiation of properties that represent the articulation of body member through a joint. The control system also manages the automatic calculation of data values from the received data, such as the size of the body members and angles formed between them.

A data instantiation example to describe a left upper limb with down gesture for the person in Fig. 4.9, would include: (i) classification of joint instances (see Fig. 4.6); (ii) `partOf_directly` property instantiations (see Fig. 4.4); (iii) joint positioning data.

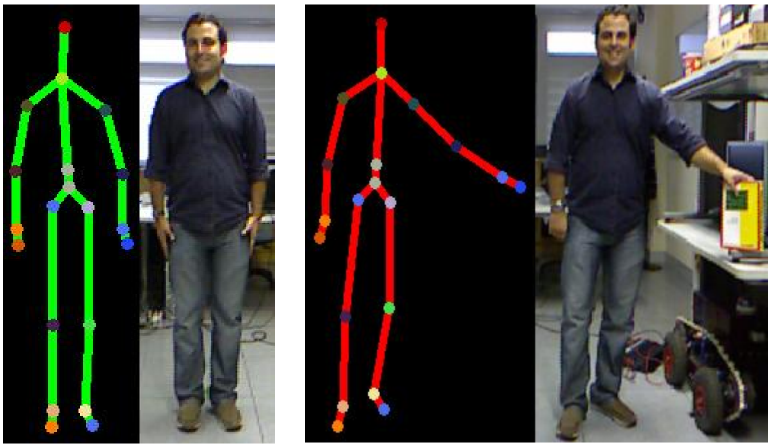


Figure 4.9: Gesture instantiation and action example

4.7.2 Activity recognition example: Touching a product

Activity recognition usually requires composition of simple activities along the time. Therefore temporal analysis is required in order to recognize complex activities (Holte et al., 2008). Our ontology model is expressive enough to represent the temporal dimension of the activities. The representation capabilities resulting from the combined use of Kinect™ and the ontology-based model offer simple but very expressive tools to detect interesting activities for a market research confection.

Relevant activities for current market researches may be: stand in front of, look at, point at and touch a product. Recognition of simple interactions between different body members and objects regarded as context data can be detected finding the spatial relationship between these elements. The process becomes more robust if the object includes sensors (e.g. RFID and accelerometer) able to provide different kinds of features –id, location and kinematic state.

In order to demonstrate the expressiveness of our representation, a syntactically relaxed nRQL –the query language of the RACER reasoner– rule is presented. Rule's variables are denoted with a question mark at the beginning of their names (?), variables belonging to the RCC substrate are labeled adding a star (?*), concept types start with a hash (#) and RCC-8 relationships are labeled with a colon (:). To the existing namespaces, tracking entities (#!tren:), scene objects (#!scob:) and activities (#!actv:), a new one is added to group all the specific information related to market researches (#!mkrs:). The syntax of nRQL has

```
1. (firerule
2.   (and
3.       (?currentFrame #!tren:CurrentFrame)
4.       (?hand #!tren:Hand)
5.       (?product #!mkrs:Product)
6.       (?person #!scob:Person)
7.       (?staff #!mkrs:Staff)
8.       (?product (> #!tren:acceleration 0))
9.       (not (?*product ?*hand :dc))
10.      (?hand ?person #!tren:componentOf)
11.      (?person ?place #!scob:placedIn)
12.      (not (?person ?staff #!scob:memberOf))
13.(
14.      (instance (new-ind ?touchingAct) #!actv:Touching)
15.      (related (?touchingAct ?currentFrame #!tren:isValidInBegin))
16.      (related (?touchingAct "unknown_frame" #!tren:isValidInEnd))
17.      (related (?touchingAct ?place #!scob:placedIn))
18.      (related (?product ?touchingAct #!actv:passivelyInvolvedIn))
19.      (related (?hand ?touchingAct #!actv:activelyInvolvedIn)))
20.)
```

Figure 4.10: Rule to exemplify expressiveness

been slightly simplified to make them more readable. The following rule detects touching activities between people and sensorized objects.

First, different variables that act along the rule are declared (3-7). The rule checks if the object involved in the situation is currently moving (8). This statement can also be used as a trigger of the rule. Afterwards, the rule checks if there is a spatial relationships between the moving `Product` and a `Hand` (9). The place of the person is assessed in (10-11). Finally, to discriminate between clients and employees, the rule considers if the person involved in the action is member of the staff (12). This identifying capability is referred in future work. If the antecedent conditions are satisfied, the consequent is applied. The consequent creates a `Touching` activity (14) with a known beginning (15) and an unknown ending (16). The spatial location of the activity is bounded by the location of the person who perform the activity (17). `passivelyInvolvedIn` and `activelyInvolvedIn` relationships among the new activity with the passive object (18) and the active subject (19) are also stated in the consequent. The resulting activity has been defined according to spatio-temporal criteria and part-based relationships.

The `Touching` activity is candidate to be classified as a subactivity of `Shopping`. To recognize the `Shopping` activity it is required to recognize a sequence of subactivities (e.g. touching the product, trying the product, interacting with the staff, paying the product) where the same active subjects and passive objects are involved in the same place and time. For the sake of simplicity a rule which only recognizes the spatial dimension of a `Touching` and a `Paying` activity is showed in Fig. 4.11.

At the beginning of the antecedent a set of variables are declared (3-6). Then, the same objects, subjects and places are identified in the subactivities (7-12). Finally, the starting and ending timestamps of the activities sequence are retrieved (13-14). The consequent

```
1. (firerule
2.   (and
3.       (?touching #!actv:Touching)
4.       (?paying #!mkrs:Paying)
5.       (?person #!scob:Person)
6.       (?product #!mkrs:Product)
7.       (?product ?touching #!actv:passivelyInvolvedIn)
8.       (?product ?paying #!actv:passivelyInvolvedIn)
9.       (?person ?touching #!actv:activelyInvolvedIn)
10.      (?person ?paying #!actv:activelyInvolvedIn)
11.      (?touching ?place #!scob:placedIn)
12.      (?paying ?place #!scob:placedIn))
13.      (?touching ?startFrame #!tren:isValidInBegin)
14.      (?paying ?endFrame #!tren:isValidInEnd)
15.(
16.      (instance (new-ind ?shoppingAct) #!mkrs:Shopping)
17.      (related (?shoppingAct ?startFrame  #!tren:isValidInBegin))
18.      (related (?shoppingAct ?endFrame  #!tren:isValidInEnd))
19.      (related (?shoppingAct ?place #!scob:placedIn))
20.      (related (?touching ?shoppingAct  #!actv:essentialSubActivityOf))
21.      (related (?paying ?shoppingAct  #!actv:essentialSubActivityOf))
22.)
```

Figure 4.11: Simplified rule to recognize shopping

creates a `Shopping` activity whose validity time interval is bounded by the starting point of the former activity and the ending point of the latter activity (16-18). The coincident place of the subactivities and the mereological properties between the subactivities and the overall activity are eventually asserted (19-21).

Crucial data is inferred from the former to the latter rule. Thanks to the interaction between the mereological and the geolocalized layers rules acquires more flexibility and the amount of relationships between concepts grows improving the completeness of the model. Imagine that the subactivities are detected in different places.

- `touchingAct placedIn GroundFloor`
- `payingAct placedIn FirstFloor`

The system can store mereological data stated to describe invariant context relationships such as:

- `GroundFloor containedIn Shop`
- `FirstFloor containedIn Shop`

In both cases, using the compositions described in Table 4.2, new relationships are inferred.

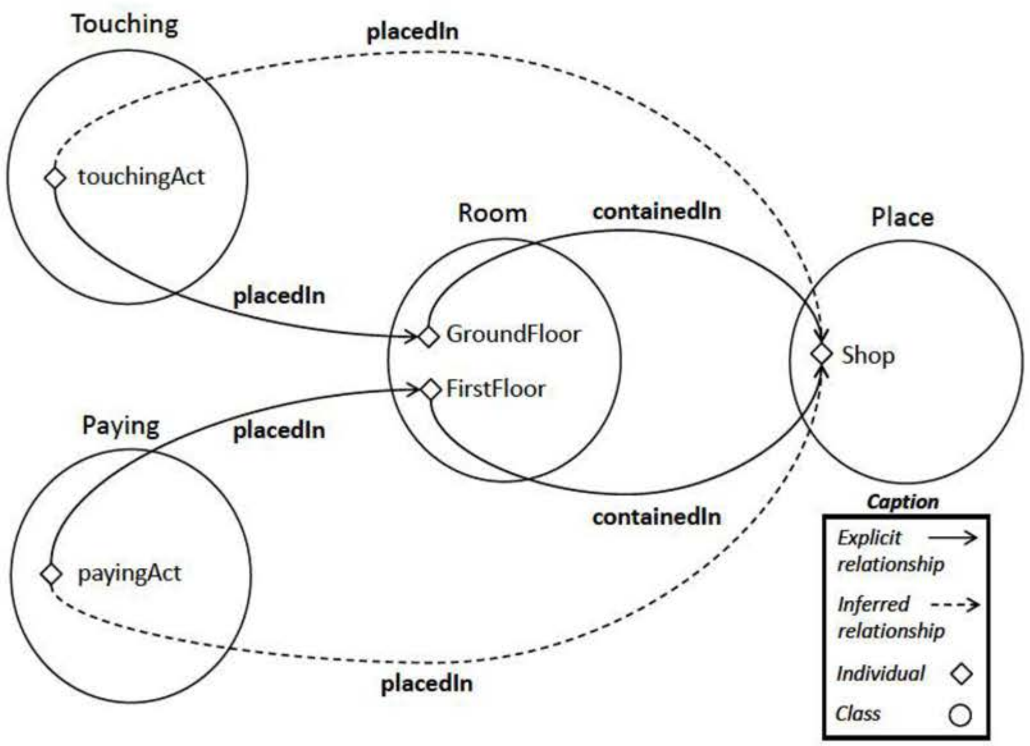


Figure 4.12: Representation of the inferred placedIn relationships

- `touchingAct placedIn Shop`
- `payingAct placedIn Shop`

Even though the activities have been detecteded in different places, the latter rule is fired because there is a common location for both activities (see Fig. 4.12). Following the reasoning, an appropriate spatial environment (`Shop`) is allocated to the overall activity (19).

4.7.3 Data propagation example: Touching a product

Many data relationships are automatically propagated from the consequent's assertions of the previous section. In the first rule (19) of the previous section, a `Hand` is declared as active subject of the `Touching` subactivity. However, in the latter rule (9-10) a previously unstated assertion includes a `Person` as active subject of this subactivity. The pattern explained in 4.5.2 justifies the propagation of activity relationships for all the parts which contains the part performing the activity. When the `Hand` was declared as an active subject, the objects containing it were also inferred as active subjects.

- `upperlimb activelyInvolvedIn touchingAct`
- `person activelyInvolvedIn touchingAct`

Data propagation enable to choose the level of granularity of the information retrieval tasks and to assess data from multiple perspectives. The following query would retrieve the interactions among the people and the upperlimbs, and the products during a campaign (it is assummed that, during a campaign, the products are located in the same place).

The query in Fig. 4.13 retrieves different levels of active subjects (`Person` and `UpperLimb`) of `Touching` activities for all the products on sale (1). Then query variables are declared (3-6). The `Product`, `Person` and `UpperLimb` of the same `Touching` activities are retrieved (7-9). From these set of activities there are only chosen those whose validity time interval is into the validity time interval of the campaign (10-12).

```
1. (retrieve (?person ?upperlimb ?product)
2.   (and
3.       (?upperLimb #!tren:UpperLimb)
4.       (?product #!mkrs:Product)
5.       (?person #!scob:Person)
6.       (?campaign #!mkrs:Campaign)
7.       (?product ?touching #!actv:passivelyInvolvedIn)
8.       (?person ?touching #!actv:activelyInvolvedIn)
9.       (?upperLimb ?touching #!actv:activelyInvolvedIn)
10.      (?touching ?touchInterval #!scob:hasInterval)
11.      (?campaign ?campInterval #!scob:hasInterval)
12.      (?touchInterval ?campInterval #!scob:intervalOf))
13.)
```

Figure 4.13: Query for different interactions during a campaign

The extracted information is helpful to answer with accountant criteria abstract questions such as: "What is the visibility of this product?" A very rough answer would be the number of people who have interact with it. The level of doubts involved in the purchase decision can be also measured if we count the number of interactions of each user with the product. An extended model able to distinguish between right and left limbs, could be used to assess the quality of the product accessibility.

Another example of propagation is the automatic assignment of subjects and objects in composed activities. The first rule of the previous section states a `Person` and a `Product` as the active subject and passive object of a `Touching` subactivity. The system automatically connects these individuals as active subject and passive object of the `shoppingAct` individual when the `touchingAct` subactivity is detected participating into a `Shopping` activity individual. This process is repeated, thanks to the composition explained in Section 4.4.1, each time a `participatesIn` property is instantiated.

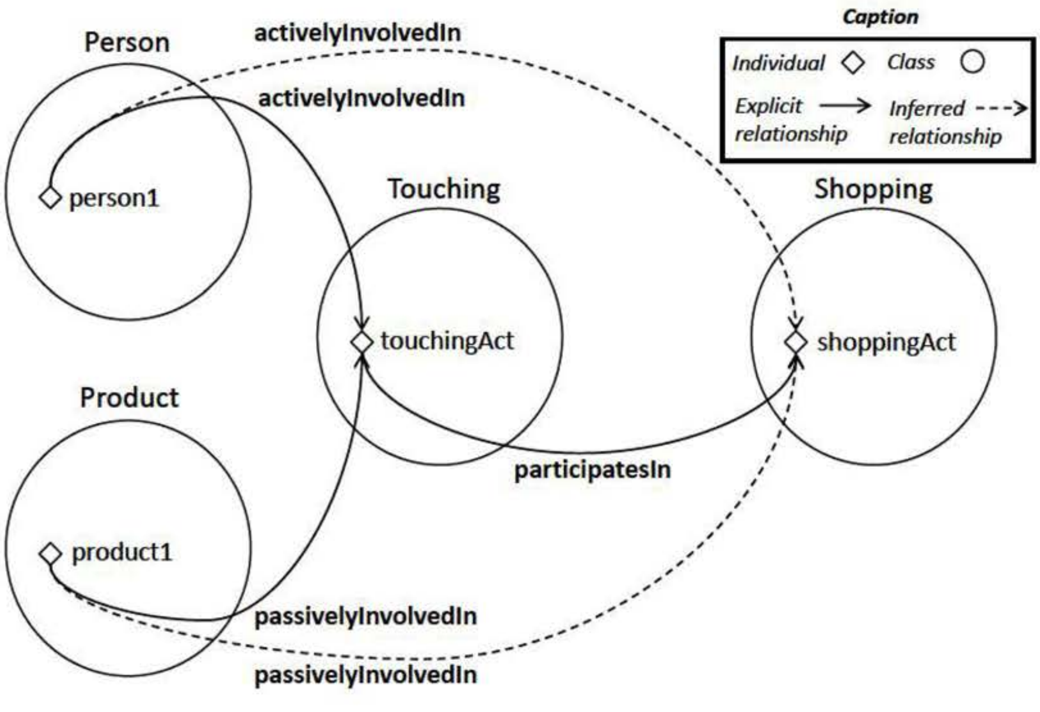


Figure 4.14: Representation of the inferred involvedIn relationships

5

Reasoning with uncertainty: A harbor surveillance prototype

Harbor surveillance is a critical and challenging part of maritime security procedures. Building a surveillance picture to support decision makers in detection of potential threats requires the integration of data and information coming from heterogeneous sources. Context plays a key role in achieving this task by providing expectations, constraints and additional information for inference about the items of interest. This chapter proposes a framework for context-based situation and threat assessment and its application to harbor surveillance. As in previous chapters, the framework uses the ontological model to formally represent input data and to classify harbor objects and basic situations by deductive reasoning according to the harbor regulations. The higher level applies Belief-based Argumentation to evaluate the threat posed by suspicious vessels. The functioning of the system is illustrated with several examples that reproduce common harbor scenarios.

5.1 Introduction

Maritime security is an area of strategic importance for the international community. As stated by Collins et al.[1], "a terrorist incident against a marine transportation system would have a disaster impact on global shipping, international trade, and the world economy in addition to the strategic military value of many ports and waterways." For that reason, one of the principal goals of strengthening maritime security is to "increase maritime domain awareness" by building a "surveillance picture as complete as possible to assess the threats and vulnerabilities in the maritime realm." In particular, harbor surveillance is a critical part of maritime security procedures because of its multiple objectives: recognition of terrorist threats, prevention of maritime and ecological accidents, detection of illegal immigration, fishing and drug trafficking, and so forth. However, it is nowadays mostly developed by human operators (Liebhaber and Feher, 2002), who have to evaluate an overwhelming amount of information. This makes it very difficult to attend to the event stream with the required

[1]Statement on transportation security before the Committee on Commerce, Science, and Transportation http://www.globalsecurity.org/security/library/congress/2004_h/040324-collins.pdf Last accessed December 2015

level of attention due to distraction, fatigue and oversight. In addition, their decisions may be strongly affected by sensor data imprecision and subjective judgment.

Next-generation harbor surveillance systems are envisioned to automatically identify potential threats with a high degree of confidence (Guard, 2007). Their objective is obtaining not only tracking information about vessels, but also an abstract picture of the situation to make informed decisions. According to the JDL data fusion model, the latter task belongs to the domain of Situation Assessment, defined as the estimation of "sets of relationships among entities and their implications for the states of the related entities." (Steinberg et al., 1999) In this domain, it requires understanding the intrinsic information provided by coastal sensors in the context determined by extrinsic factors, like harbor environment, operational regulations, traffic data and intelligence reports.

The increasing interest in higher-level information fusion has led to several proposals for context management. Detection and characterization of activities and threats require assessing the states of situational items and their relationships within a specific context (see A). This contextual information, expressed in the form of complementary knowledge or constraints, encompasses information about objects, processes, events, and relationships between them, as well as particular goals, plans, capabilities, and policies of the decision makers. Such diversity makes formal context representation a significant challenge.

Ontologies are an appropriate formalism to represent contextual and factual knowledge in higher-level fusion (Nowak, 2003), (Little and Rogova, 2009), (Kokar et al., 2009) However, ontology languages based on Description Logics, and in particular the standard ontology web language (OWL 2)[2], present several unsolved challenges when applied to Situation Assessment because: (i) they do not allow for reasoning with uncertain knowledge; (ii) they do not directly support abductive reasoning to create and validate situational hypotheses that change in time.

This chapter describes an Information Fusion system that uses contextual knowledge represented with ontologies to detect and evaluate anomalous situations. By contextual knowledge we mean knowledge about external information that completes, influences or constrains the situations or events of interests; e.g. physical characteristics of the environment such as terrain or weather, or knowledge about the expected behavior of the objects. The system is arranged in two processing levels. Firstly, the system applies deductive and rule-based reasoning to extend tracking data and to classify objects according to their features. Secondly, the Belief-Argumentation System (BAS) –a logic-based paradigm for abductive reasoning (Rogova et al., 2006) –is used in combination with the Transferable Belief Model (TBM) (Smets and Kennes, 1994) to determine the threat level of situations involving objects which are not non-compliant to a normality model. A prototype implementation of this system adapted to the harbor surveillance problem is available for experimental evaluation in a public repository [3].

To the best of our knowledge, this is the first attempt to combine ontologies and TBM based uncertain reasoning to implement multi-level information fusion. Similar approaches in the literature have focused on alternative probabilistic models (see C). Ontologies facilitate

[2] W3C OWL Working Group, OWL 2 Web Ontology Language Document Overview. http://www.w3.org/TR/owl2-overview Last accessed December 2015

[3] Harbor simulator project. https://github.com/meditos/HarborSimulator Last accessed December 2015

the creation of a computable model representing complex situational context (problem entities, scenario geometry, spatial relationships, etc.), since they can be formally encoded in a logic-based expressive language. The examples show that this integrated approach reduces the number of false alarms with respect to purely ontological proposals through quantifying the threat level.

5.2 Context and Ontologies in Information Fusion

This chapter follows the definitions given in (Powell et al., 2006), (Kandefer and Shapiro, 2008) which refers that context is any external knowledge that is useful or influences the fusion processes, including background knowledge (e.g. tactics, doctrine), situation-specific knowledge (e.g. terrain), existing reports and databases, and so forth. In this case we create a model of the scenario and use background, situational and expert knowledge to drive the high-level fusion process. Specific contents of the context model for the harbor surveillance problem are described in section 5.3. Context can thus be used to explain observations, to define hypotheses, to identify areas of interest to focus new data collection, to refine ambiguous estimations, and to provide for interrelationship between different fusion levels (Sycara et al., 2009), (Gómez-Romero et al., 2010).

The complex uncertain harbor surveillance scenario calls for a hybrid context representation combining ontology and logic-based models enriched by uncertainty consideration. We propose a fusion system in which the description of the domain entities, such as vessel types and harbor areas, the relations between them, and applicable regulations are modeled as a certain set of ontological concepts, relations, instances, axioms and rules. Deductive reasoning is applied to detect inconsistencies between the situations obtained as a dynamic instantiation of the scene model and the situational patterns defined in the normalcy model. Normalcy rules are local to a navigational context, which depends in most cases on the geographical situation of the vessel (as in (Snidaro et al., 2015)). Inconsistencies denote abnormal situations that may indicate a potential threat. A probabilistic reasoning process is then triggered to investigate whether these inconsistencies are the result of insufficient quality of observations, contextual knowledge, and fusion processes, or the result of the change of context; e.g. discovered potential or imminent threat.

5.2.1 Ontologies, Logic and Uncertainty in Higher-Level Fusion

During the last decade, several approaches using ontologies have emerged in the higher level Information Fusion research area. The SAW Core ontology represents general concepts used in situational awareness (Matheus et al., 2005). It was used as a meta-model in (Baumgartner et al., 2010), which applied deductive reasoning for Situation Assessment in a traffic-management scenario. The Situation Theory Ontology (STO) has been recently proposed as a formal upper model to represent the abstract concepts involved in Situation Awareness under the semantics of Barwise and Perry's situation theory (Kokar et al., 2009). In the harbor domain, ontologies have been also used to represent a priori and contextual information. In (Roy and Davenport, 2010), a MDO (Maritime Domain Ontology) was created to automatically classify vessels and situations from perceived situations by applying deductive reasoning. Similarly, in (Van den Broek et al., 2011) the authors showed that ontologies are useful to capture ancillary knowledge on the elements of the application domain and behavioral patterns.

In these works, situation recognition is mostly achieved by instance classification as follows. Context models include ontological descriptions of categories of entities and situations. When a new object is created or its property values are modified, a deductive reasoning process finds matches to these descriptions and determines the class or classes to which the new instance belongs. However, this procedure is insufficient in complex Situation Assessment problems, because there is an inherent uncertainty in this process that is not considered, and more than one hypothesis may explain the current situation, but only one is generated. In general, ontologies are not suited for abductive inference and reasoning under uncertainty (Lukasiewicz and Straccia, 2008). In addition, they are not particularly effective to represent perdurants; i.e., entities that change in time. This requires the creation of artificial representational patterns (Gómez-Romero et al., 2011b) or the use of non-standard extensions to the standard languages (Motik, 2012). As introduced in the beggining of this chapter, an extension of the typical deductive reasoning with probabilistic abductive reasoning.

One of the most common approaches to incorporate uncertain, unreliable and imprecise knowledge to Web ontologies is PR-OWL 2, an extension of the OWL language with Bayesian probability theory (Carvalho et al., 2013), that has been illustrated with examples on higher-level fusion in the maritime domain (Carvalho et al., 2011), (Laskey et al., 2011). PR-OWL 2 represents factual and contextual knowledge in terms of instances and properties with associated uncertainty. Currently, there is not any available reasoner that entirely supports this language. However, the resulting ontologies can be transformed into a probabilistic network and processed with the UnBBayes [4]. tool. The main difference regarding this proposal is that the probabilistic formalism is do not embedded into the ontology. Instead, ontologies are used as a unified representation in a deductive layer to extend data with available knowledge, and delegate threat assessment tasks to the upper layer implementing the BAS-based reasoning process. This adds more flexibility to the system and reduces the computational cost that is often associated to ontology-based reasoning, which may make the application unusable under real-time restrictions.

A related proposal is presented in (Snidaro et al., 2015). The authors use Markov Logic Networks (MLNs) to represent uncertain context knowledge and automatically detect anomalies in the maritime domain. MLNs combine the expressiveness of first order logic and the uncertainty management of Markov Networks, thus providing a very intuitive and powerful knowledge framework. The treatment of context information is very similar as in this approach, since it is conveniently integrated into the representation and exploited to properly interpret the available data. The paper does not study in detail the possible effects of the semi-decidability of first order logic, which may be a drawback compared to decidable Description Logics ontologies. Besides, they assume that information is already available and expressed in a symbolic form, as in (Carvalho et al., 2011), (Laskey et al., 2011). It is not clear how the raw sensor data is incorporated into the logic model, a problem that is explicitly tackled in this proposal.

[4]UnBBayes web page. `http://unbbayes.sourceforge.net/` Last accessed December 2015

Building a common framework for the evaluation of probabilistic higher-level Information Fusion systems is a research topic that has received considerable attention recently. The Evaluation of Techniques for Uncertainty Representation Working Group[5]. (ETURWG), hosted by the International Society of Information Fusion (ISIF), is an ongoing initiative purposely formed in 2011 to address this problem. The URREF (Uncertainty Representation and Reasoning Evaluation Framework) ontology is an initial proposal towards the formal description of the concepts involved in a probabilistic fusion system and the applicable comparison criteria (Costa et al., 2012). Nevertheless, the current state of this proposal makes it still unfeasible to carry out a detailed comparison among different systems.

5.3 Sensorial and Contextual Information in the Harbor Surveillance Scenario

Surveillance picture formation in the harbor scenario is the result of a multi-level fusion process, which includes:

- Data acquisition from heterogeneous sources about single objects.
- Object tracking to integrate sensor data and obtain the tracks (location, kinematics, identification) representing all objects present in the scene.
- Object property estimation for object categorization.
- Utilization of context knowledge about expected object properties, identification, and behavior to classify objects and to infer basic relationships and situations.
- Matching expected behavior provided by the context of entities to the observed situation in order to detect a possible anomaly as an initial step towards scene recognition.
- Abductive reasoning to explain inconsistency, to detect possibly threat to alert an operator, and to improve the overall functioning of the system and the knowledge base.

Input data encompasses hard and soft sources, ranging from sensor measurements to intelligence reports. Sensor data is automatically acquired by primary coastal sensors or cooperatively emitted by ships. The main primary-sensor technology for object detection and location in the harbor is the coastal radar, which does not require cooperative equipment installed onboard of ships. Therefore, the low level input of the system is either raw position measurements (in a centralized architecture) or fused estimates obtained by a processing node (in a decentralized architecture). In both cases, the fusion node involves three basic functions: (i) data alignment or common referencing involving coordinate or units transformations, uncertainty normalization, and inter-sensor alignment; (ii) data association to determine to which measurements are associated to which entities; (iii) state estimation involving the computation of entity attributes at Level 1 –e.g. location, velocity, and other

[5]Evaluation of Techniques for Uncertainty Representation Working Group web page. http://eturwg.c4i.gmu.edu/ Last accessed December 2015

classification attributes such as size or category. Ships also emit identification data according to IMO (International Maritime Organization) security protocols, mainly through the Automatic Identification System (AIS). The AIS system broadcasts basic data obtained by the available navigation equipment (identification, position, course, and speed) together with extended data (intended route, cargo description, etc.). Other relevant data sources are the Vessel Traffic Systems (VTS), which frequently collect all available inputs in an integrated tracking image (Guerriero et al., 2008) and the Port Traffic Management Systems (PTMS) (Seibert et al., 2006).

For practical purposes this chapter assumes a preexisting decentralized tracking schema with a working fusion node located after a set of single-source target tracking systems. This schema provides vessel tracks with reasonable accuracy already available to be processed. In general, a decentralized solution is more realistic in the maritime surveillance scenario, since it allows using the available tracking systems and taking into account the different data types and update rates of AIS and VTS. The tracking sub-system could also benefit from the available context information. For instance, ships trajectories might be constrained to follow the assigned channels according to deep draught category and water depth. A dynamic model for vessel track prediction can be used to incorporate this knowledge into the tracker. Table 1 describes some elements of static –a priori, or configuration data– and dynamic –a posteriori, or information inferred at the same time as sensor data is obtained– contextual information of interest.

5.4 Reasoning Schema

Figure 5.1 depicts the processing layers for dynamic surveillance picture formation. Firstly, tracking and object identification data is fused to obtain track features and used to update the scene model. In this layer, the scene ontology defines the concepts and relations of the surveillance problem. Concepts are represented by ontology classes, whereas relations are represented by ontology properties. Accordingly, tracked entities are asserted into the model as class instances. Spatial relations among vessels and other elements in the scene (harbor channels, mooring positions, constrained areas, etc.) are also calculated at this point. Purposely, topological reasoning is performed to detect and update qualitative topological relations aomng the scene elements. This procedure is explained in Section 5.5.2.

Once sensor information is symbolically represented in the scene ontology, a classification procedure is performed to determine the type of the vessels according to their features and their topological properties. Here we use 'type' in a wide sense, since the outputs of this process are statements describing vessels by their features (size, flag, function, etc.) and behaviors (stopped, exceeding channel speed, too close to other object, etc.) Next, contextual information together with all available transient information is used to classify the situations for each object or group of objects. That is, the behavior of estimated situational items is compared with the corresponding expected behavior in the context under consideration. This procedure is explained in Section 5.5.2.

When the estimated situational items are different from expected, it is necessary to understand the source of this discrepancy. The difference can be attributed to the poor quality of the observations and the limitations of the tracking process (e.g., sensor noise,

Table 5.1: Contextual information sources in the harbor surveillance scenario

Static context knowledge
Ships characteristics and behavior restrictions, such as speed, functions, etc.
Geographic knowledge with environmental maps: harbor configuration, coastline, currents, channel navigability, restrictions, etc.
Navigation knowledge describing how vessels maneuver as they progress along shipping channels, meet other vessels, and encounter different weather.
Sensor characteristics: areas of poor radar coverage, presence of clutter regions.
Operational rules about coordinated motion of several vessels; e.g., mandatory use of tug boats to escort the cargo ships from the inner port entrance until the final mooring position.
Allowed proximity to other vessels, protocols for collision avoidance, and rules of precedence.
Information on intended vessel trajectory: sailing plan or pre-established route, estimated times of arrival, etc.
Dynamic context knowledge
Environmental parameters: modifications of channel navigation restrictions, allowed areas depending on time of day, etc.
Sea conditions, ice.

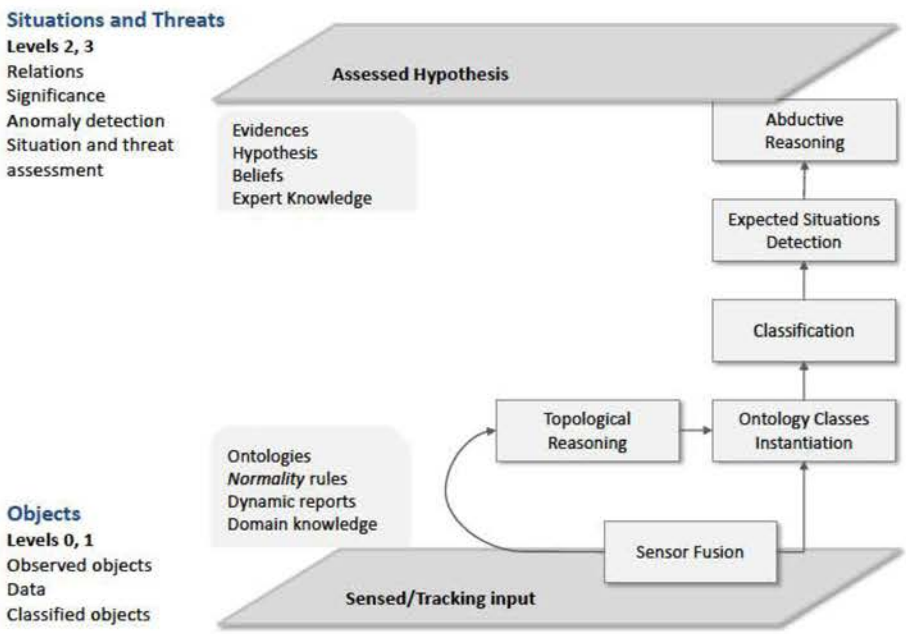


Figure 5.1: Workflow of the multi-level fusion system in the harbor domain

bad resolution, continuity problems and association errors), the use of imperfect or erroneous knowledge, or the existence of a real threat. To make a distinction, the system triggers the abductive reasoning process aimed at explaining the source of inconsistency and assessing the possible threats (Section 5.5.3).

5.5 Situation Detection

5.5.1 Detection of Normal Situations

The procedure of deviation detection from the normalcy model is performed in several steps, as explained before. This section presents a context model for harbor operations representation based on ontologies, the reasoning procedures that are applied for vessel classification and expected situation detection based on rules expressed in the Semantic Web Rule Language[6] (SWRL) rules –the de facto standard for rule-based reasoning with OWL ontologies and the procedure to encode and instantiate the topological predicates.

5.5.2 Representation of Vessel Characteristics and Harbor Areas

Vessels are represented as instances of the ontological model. Most vessel properties, such as speed and position are transient; i.e., they change during the existence of the vessel object. To represent these changes in the ontology, it is necessary to associate vessel snapshots to vessel instances (see 2.2). More details of the ontological representation of these entities can be found in (García et al., 2011). For the sake of simplicity, in the remaining sections we will assume that transient properties are directly assigned to vessel instances without using snapshots.

Geographic knowledge of the harbor can be represented at different levels of granularity. Typically, there are two different areas in a harbor: the land area including inner water, which is the port, and the outer water area, which is called the road. No ship can enter in the port without the permission of Harbor Master's Office after reporting requested details such as identification code, nationality, length, draught, cargo, and so forth. The anchorage is the designated area on water where ships wait for the entrance to the port. Inside the port, different facilities used for ship mooring and berthing can be identify. Harbor authorities define navigation areas for different categories of vessels, e.g. separated channels for small power-driven vessels, big power-driven vessels and nonpower vessels. In addition, navigation near to certain facilities may be restricted or even forbidden.

Figure 5.2 shows an excerpt of the ontology representing a scenario with a vessel, two navigation channels and a restricted facility. At a basic level, zones are manually described by means of the global coordinates of their delimiting polygon. Vessel location, in turn, is represented with a punctual position estimation resulting from fusing radar and AIS information. At a higher level of abstraction, vessel relative positions with respect to zones, as well as zone relative positions with respect to other zones, are represented through qualitative

[6]SWRL: Semantic Web Rule Language Combining OWL and RuleML. http://www.daml.org/2003/11/swrl Last accessed December 2015

spatial properties that relate different entities, rather than entities and data values. Additional abstract spatial relations are defined inside the scene model; e.g., *close to* (proximity between vessels or vessels and facilities) and *aligned to* (angle between vessels and channel navigation directions).

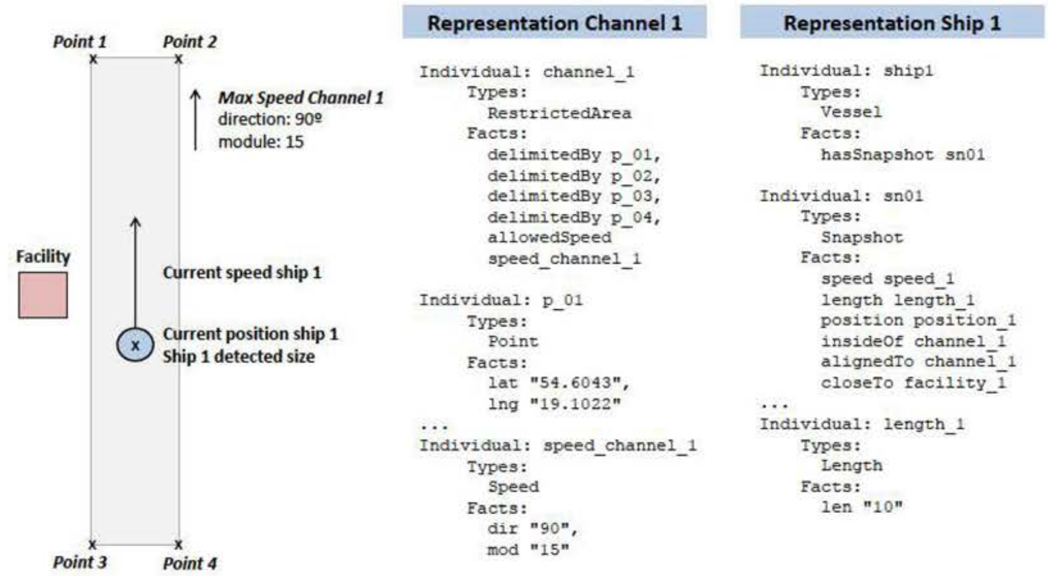


Figure 5.2: Representation of harbor zones, facilities and vessels

These relations require some geometrical calculations to be instantiated. For example, it is necessary to determine if the distance between two entities is less than a threshold in order to instantiate the property *close to*. This process is performed by the topological reasoning module. For the implementation of this module, we have used the OpenGIS standard and the Java Topology Suite[7], a programming library to calculate geometrical relations between positioned entities. It is important to notice that in a first 'brute force' approach, topological relations are calculated between each pair of entities when one of them is updated. This requires a considerable amount of computations, and necessarily calls for the implementation of optimized geometric models able to segment the space in influence zones, in such a way that the number of property calculations would be dramatically reduced (Serrano et al., 2011). In order to express the topological relations the Region Connection Calculus (RCC) has been choosen as topological formalism. RCC is a logic theory for qualitative spatial knowledge representation and reasoning (Renz, 2002). RCC semantics cannot be fully represented with ontologies (Grütter et al., 2008), but typical reasoning engines provide support for them through an extended processing layer (see B.1).

[7]Java Topology Suite web page. `http://www.vividsolutions.com/jts/JTSHome.htm` Last December 2015

5.5.3 Reasoning for Vessel Classification and Expected Situation Detection

Ontologies provide strong support for deductive reasoning, defined as an automatic procedure aimed at inferring new implicit axioms that have not been represented but are entailed by the explicit axioms. Basic ontological reasoning is concerned with the inference of subsumption axioms (i.e., determining the implicit taxonomy according to asserted classes features) and instance membership axioms (i.e., determining the type of an instance according to asserted classes and individual features). Reasoning algorithms are implemented by inference engines like Pellet (Sirin et al., 2007), the one used in the prototype presented in this chapter.

Instance membership inference is used to classify vessel instances. For example, we can define a class for small boats to include all ships that have a length less than or equal to 15 meters. To do so, an equivalence axiom is used. If a new vessel instance is asserted into the ontology with length 10 meters, or the length property value of an existing vessel instance changes to a compliant value, the vessel is automatically inferred as a member of the small boat class. Accordingly, the boat detected in Figure 2 would be classified as a small boat. We show a few example class definitions to classify vessels in Section 5.6.2.

Context knowledge is included not only to classify vessel types, but also to represent and reason with the harbor regulations that determine whether a vessel is exhibiting a normal behavior. This is the normalcy model of the harbor: a collection of rules that are used to classify vessel behavior as compliant to the navigation rules or not. The model characterizes predictable behaviors according to harbor rules, rather than describing the features of an attack, since the complete enumeration of such unexpected events is, by definition, incomplete. The open world assumption, which stands when reasoning with ontologies, favors this kind of representation. This assumption states that, by principle, the set of axioms in the knowledge base is not complete, and therefore, new knowledge cannot be inferred inductively. In practice, that means that an axiom that is not entailed by the model is not inferred as false, but as unknown. For instance, according to the previous example, if a vessel instance has a length larger to 15 meters, trivially the reasoner would not conclude that it is a small vessel. Nevertheless, it would not either decide that it is not, because there is not enough knowledge to confirm the latter inference. If other assertions lead to classify this instance as a small vessel, then the ontology would be inconsistent.

The normalcy model includes not only the description of "good", expected, behaviors (positive information/vessels *musts*) but also the description of situations that obviously break the harbor rules (negative information/vessel *prohibitions*). The former are useful to directly include harbor rules into the model (compliance conditions), whereas the latter allows the inference system to check the existence of predefined suspicious or threatening behaviors (violation conditions). This is made to improve system performance, because selected situations are directly classified as abnormal, and to facilitate modeling, because in some cases it is easier to express a harbor navigation rule by presenting the cases that are not compliant to it, instead of those that are. In any case, as mentioned, vessel behavior can be classified only if there is enough evidence according to its properties. Among classified behaviors, we have vessels: (i) compliant to harbor rules, (ii) not compliant to harbor rules; (iii) compliant to some harbor rules and not compliant to some harbor rules. In Section 5.6.2 we show an excerpt of the hierarchy of expected situations of the example.

Harbor rules are expressed in the normalcy model with class inclusion axioms and rules.

Class inclusion axioms can be used to describe under which circumstances a vessel is included in the compliant/not compliant behavior classes, in a similar way as it is done for vessel classification from properties. More interestingly, SWRL rules generalize class inclusion axioms by allowing the use of bounded variables in the antecedent and the consequent of the rule. SWRL supports deductive inference with OWL ontologies under certain safety restrictions to guarantee decidability of the representation (Motik et al., 2005). Essentially, the safety restrictions limit the use of variables in rules to pre-existing named entities. This forbids adding new factual knowledge (i.e., creating new instances) during reasoning, which also implies that scene interpretation through abductive reasoning is not directly supported. We use SWRL rules to classify vessels behavior according to the harbor navigation rules. This gives an initial description of the scene in terms of the expected situations detected. For example, we can define a rule to state that a vessel aligned to its enclosing navigation area is satisfying the navigation direction requirements of the harbor. Note that at huge quantity of harbor restrictions can be easily modeled by using the concepts defined in the ontology as an abstract vocabulary.

If we consider the processing architecture shown in Figure 5.1 and the ship depicted in Figure 5.2, the workflow for object classification and situation deduction is as follows. First, the topological reasoning module detects that the ship is inside a navigation channel, and consequently instantiates the property "inside of". The topological module also detects that the ship is aligned to the enclosing zone(s) and instantiates the relation "aligned to". Next, the corresponding rule is fired and the behavior of the ship is automatically classified as compliant to the harbor rules. If it were non-compliant, this information would be provided to the uncertainty module for the construction of threat beliefs.

5.5.4 Hypothesis Evaluation for Situation and Threat Assessment

In previous works, authors have considered two complementary dimensions of context knowledge that are relevant to characterize an entity X (Steinberg and Rogova, 2008), (Gong, 2005): in the Context of X (CO) and Context for X (CF). CO encompasses the sets of situations or events that form the environment itself; e.g., the context of normal operations in the harbor (all the rules defined by port authorities are obeyed). It defines expectations about the entities, and may be used to predict observations or to trigger abductive reasoning in case of deviations. On the other hand, CF defines the items externally related to and referenced by X. In the harbor surveillance domain, it includes extraneous characteristics such as the weather, time of day, harbor geometry, buildings, etc.

Knowledge of the harbor describing objects, their properties, and behavior is used to define the expected surveillance picture. Detected deviation from the normal surveillance picture may have several possible explanations, or underlying causes. It can be the result of insufficient quality of Level 1 estimations; e.g. inaccurate and unreliable tracking. It can be also caused by utilization of the wrong environmental conditions –wrong CF– in processing sensor information (e.g. failure to correctly take into account fog in computing sensor reliability), or by employment of a certain type of sensors (e.g. a night vision sensor during day time) leading to incorrect classification of the objects and their behavior (such as noisy estimation of heading, wrong vessel category, etc.) The deviation may be as well a consequence of poorly estimated or described characteristics of the current situation, or underlying change in the current situation –change in CF. These cases can happen as a result of possible terrorist or pirate threat leading to the change of the global harbor procedures and constraints. Therefore, it is important to detect and understand the cause of anomaly to alert the operator and trigger an appropriate response. This abductive process of inferring the cause as an explanation of the effect encompasses the creation of hypotheses to explain the state of the world, the computation of the credibility of these hypotheses, and the selection of the most credible hypotheses (Josephson, 1990), (Thagard and Shelley, 1997). The hypothesis evaluation process needs to consider: (1) to what extent the selected hypothesis is better than the alternatives; (2) how credible the hypothesis is, without regarding the alternatives; (3) the quality of incoming data and information on objects and their behavior, which requires explanations (see C.4).

In the harbor surveillance problem, the task of hypothesis evaluation is based on the analysis of:

- Vessel features (speed, direction, type, flag, etc.).
- Spatio-temporal relations between the vessels and the boats in general and the harbor areas.
- Beliefs assigned to assumptions based on the observed spatio-temporal relations and correspondence of the boat behavior to rules and regulations as well as quality of transient information.

For example, we can consider the following argument *pro* hypothesis "threat" from a boat: *a boat is too close to a big vessel 'and' the big vessel is a tanker 'and' the boat is increasing its speed.* Thus this argument is a conjunction of three assumptions:

(1) A boat is too close to a big vessel

(2) The boat is increasing its speed

(3) The big vessel is a tanker

In this use case belief measures can be modeled as functions of boat dynamics (increased speed), type of the vessel (a tanker), and the relation "close" between the boat and a tanker. Thus, the belief in "too close" can be measured as a function of the difference between the observed and allowed distance between the tanker and the boat; and the accuracy and reliability of the distance observed. The next section illustrates in detail the utilization of this approach.

5.6 Example: Traffic Surveillance in a Harbor Scenario

5.6.1 Description of the Scenario

The case study considers a frame of discernment with two hypotheses $\Theta=\{\theta_1,\theta_2\}$ corresponding to "threat" and "no threat", which are evaluated for each entity in the scenario. It has been built from available descriptions of regular operations in real harbors and the associated traffic regulations of daily activities[8]. This frame entails a simplification of the complete procedure explained in C.4, because the number of hypothesis is reduced and we do not consider the hypothesis selection procedure. Hence, it is not referred as abduction, but just as threat detection.

Context information includes the geometry of the harbor navigation channels, the rules and restrictions related to the normal navigation patterns and the special navigation procedures allowed in these channels. In particular, it includes special navigation procedures within inner harbor requiring the use of towing boats for certain size and cargo category of commercial vessels. The scenario considers four different kinds of channels (Figure 5.3): special container channels (SC) for ships with special cargos that must be towed; harbor ship channels (HS) for serving boats; general cargo ship channels (GC) used for transportation purposes; and small boats channels (SB), used by recreational boats and ferries. Each channel is denoted by two letters representing its type, and one or more letters to specify the allowed navigation directions (N, S, E, W). In addition, the harbor also includes a restricted navigation area next to the SCE1 channel in the surroundings of a liquid fuel terminal (LFT). The operation rules considered are described in Table 5.2.

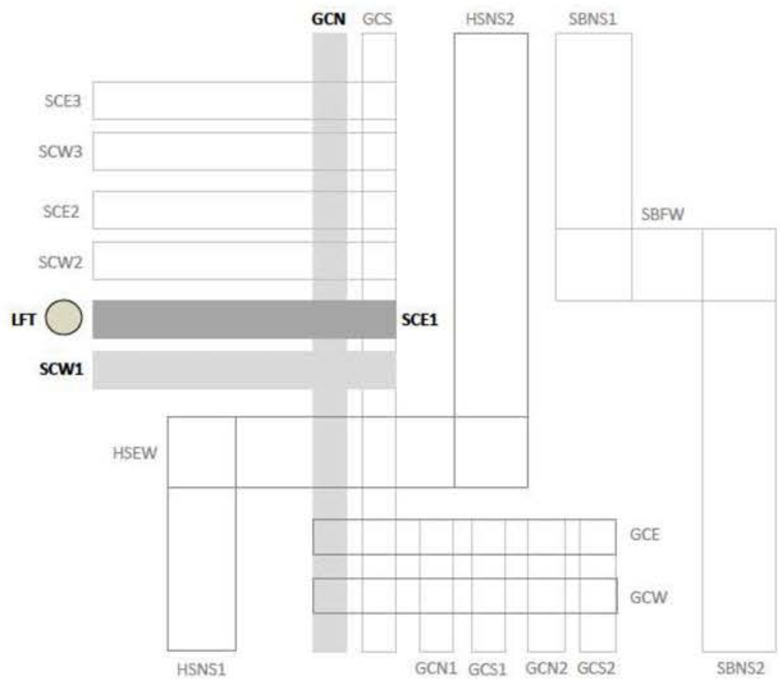


Figure 5.3: Scenario harbor zones (zones used in the example are highlighted)

[8]Port of Gdansk web page http://www.portgdansk.pl/events/vts-gulf-of-gdansk Last accessed December 2015

Table 5.2: Example harbor regulations

<table>
<tr><th colspan="2">Speed Limits</th></tr>
<tr><td>General Cargo Channel</td><td>15 knots for all ships</td></tr>
<tr><td>Special Containers Channel</td><td>10 knots for all ships</td></tr>
<tr><td>Small Boats channel</td><td>12 knots for all ships</td></tr>
<tr><td>Harbor Ships Channel</td><td>20 knots for surveillance ships
15 knots for other ships</td></tr>
<tr><th colspan="2">Alignment</th></tr>
<tr><td>North (N)</td><td>90°</td></tr>
<tr><td>South (S)</td><td>270°</td></tr>
<tr><td>East (E)</td><td>0°</td></tr>
<tr><td>West (W)</td><td>180°</td></tr>
<tr><td colspan="2">Ships in crossing areas should be aligned at least with one of the directions</td></tr>
<tr><th colspan="2">Towage</th></tr>
<tr><td colspan="2">Ships of 70m and more in length, carrying dangerous cargo, shall be obliged to use tug service while entering the port (from the road to mooring position at the port), while leaving the port (from unmooring to the road), and at every change of berth within the port area. Specifically, 2 tugs are required for:
• Ships of length over 170m
• Ships and floating facilities without propulsion of length over 130m
• Special ships of length over 150m
Tug boats must be into the towing perimeter of the assisted ships and aligned while towing</td></tr>
</table>

Table 5.3: Definition of ontology classes to classify vessels according to detected properties

Properties	Vessel type
VESSEL *AND* (LENGTH *SOME* (LENGTH *AND* (LEN *SOME* DOUBLE[<= 15])))	SMALLBOAT
VESSEL *AND* (LENGTH *SOME* (LENGTH *AND* (LEN *SOME* DOUBLE[>=170])))	LARGEPOWERDRIVENVESSEL

5.6.2 Context, Assumptions, Arguments and Beliefs

A simple harbor specific OWL ontology has been developed with the Protégé 4[9] editor including the classes, properties, axioms and rules necessary for the example. It comprises 32 classes (16 of them for vessel classification purposes) and 30 properties (10 of them are topological predicates). Table 5.3 shows two class definitions used for classification of small boats and large power driven vessels. In addition, the ontology includes definitions for expected situations, corresponding to behaviors compliant and non-compliant to navigation rules. Figure 5.4 shows an excerpt of the taxonomy of consistent behaviors and safety violations.

Table 5.4 describes the rules that define the assumptions and the arguments used to check the normalcy model of the scenario. Some of these rules are based on predicates that are instantiated by the topological reasoning module. For example, in the rule referred in the second row of the table, the normalcy model classifies a ship as a SPEEDVIOLATION instance since it is faster than the maximum speed allowed for the area in which it is currently moving. In the experiments shown, four types of arguments are considered: violations of speed limit, violations of navigation direction, incorrect towing operations and violation of protected facilities. For the sake of simplicity, all the arguments considered in the example are pro hypothesis "threat". The detected abnormal behavior at time t_i triggers reasoning aimed at explaining inconsistency and deciding whether this inconsistency points to a threatening behavior.

In the case of speed limit, the argument pro hypothesis "threat" (*Arg1*) is based on detected speed violation and represented by a conjunction of two uncertain assumptions (*A11* and *A12*) and a proposition (*P1*):

$Arg1 = A11 \wedge A12 \wedge P1$

where:

A11: The boat is inside area X

A12: The speed of the vessel is greater than Y

P1: The speed limit in X is Y

Beliefs in the arguments are calculated by combining the belief that the assumptions are true. We consider the hypotheses $\Omega = \{T^I, F^I\}$, where T^I is a hypothesis that assumption I

[9] Protégé web page. http://protege.stanford.edu Last accessed December 2015

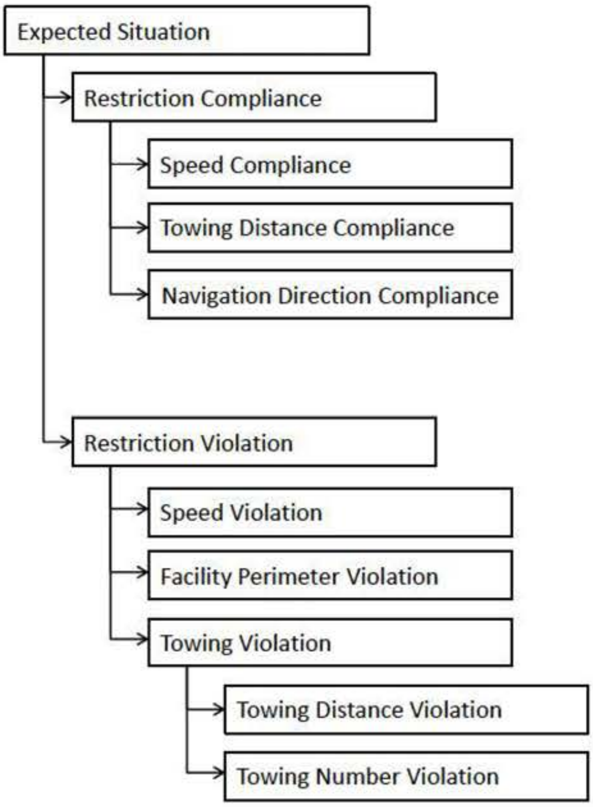


Figure 5.4: Excerpt of consistent behaviors and safety violations in the ontology

Table 5.4: Definition of ontology rules to classify vessels behavior according to expected situations

Assumptions and arguments	
Ship navigation angle is compliant to the one of the including area or channel	
VESSEL(?X), ALIGNEDTO(?X,?A), INSIDEOF(?X,?A)	NAVIGATIONDIRECTIONCOMPLIANCE(?X)
Ship speed is faster than the maximum speed allowed for the including area	
AREA(?A), VESSEL(?X), SPEED(?X,?S), MOD(?S,?VS), INSIDEOF(?X,?A), MOD(?A,?VA), GREATERTHAN(?VS,?VA)	SPEEDVIOLATION(?X)
Perimeter violation of a secured facility	
DANGEROUSFACILITY(?F), VESSEL(?X), CLOSETOFACILITY(?X,?F)	FACILITYPERIMETERVIOLATION(?X)
Large power driven vessels must be towed by a predefined number of boats	
LARGEPOWERDRIVENVESSEL(?X), CURRENTTOWINGBOATNUMBER(?X,?Z), RECOMMENDEDTOWBOATNUMBER(?X,?Y), NUM(?Y,?A), NUM(?Z,?B), NOTEQUAL(?A,?B)	TOWINGNUMBERVIOLATION(?X)
A vessel is being towed by two boats, but they are not correctly aligned for this operation	
VESSEL(?X), VESSEL(?Y), VESSEL(?Z), ISTOWEDBY(?Z,?X), ISTOWEDBY(?Z,?Y), NONALIGNEDTOTOW(?X,?Y)	TOWINGALIGNMENTVIOLATION(?X), TOWINGALIGNMENTVIOLATION(?Y)

is true and F^I is a hypothesis that assumption I is not true. The measures of belief for each assumption are modeled as follows.

- For *A11*:

$$bpa(T^{A11}) = exp(-\lambda_{A11} \cdot \mid DO - W \mid), bpa(F^{A11}) = 0, bpa(\Omega) = 1 - bpa(T^{A11}) \tag{5.1}$$

  where:

  - $DO = d_{left} + d_{right}$
  - d_{left}, d_{right} are observed distances to the left and right bound of the channel
  - W is the width of the channel

- For *A12*:

$$bpa(T^{A12}) = 1 - exp(-\lambda_{A12} \cdot \mid OV - MV \mid), bpa(F^{A12}) = 0, bpa(\Omega) = 1 - bpa(T^{A12}) \tag{5.2}$$

  where OV, MV are the observed and maximum boat speeds, respectively

  ($\lambda_I \in (0,1)$, $I = A_1 1, A_1 2$ are the scaling parameters for these *bpa*)

Tug boats and vessels must be aligned to the channels in which they are into. An alignment violation may indicate threat. For instance, for a boat moving within SCW1, the argument in support of normal operations *Arg2* is based on correct alignment, and is represented by conjunction of two uncertain assumptions (*A21*, *A22*):

Arg2= *A21* $\wedge$ *A22*

where:

A21: the boat is in SCW1

A22: the boat is going in the right direction

The belief that these assumptions are true is computed based on features representing the position and direction of the movement, which are obtained from tracking and the allowed navigation directions. For example, *bpa* for *A22* can be computed as a function of the differences between observed and allowed angles:

bpa that the boat is going in the right direction is defined as follows:

$$bpa_1(T^{A22}) = \frac{1 + cos(\sigma)}{2} \chi^1_{22}, bpa_1(F^{A22}) = 0, bpa_1(\Omega) = 1 - bpa_1(T^{A22}) \tag{5.3}$$

where σ is the angle between the observed and allowed directions *bpa* that the boat is moving in the opposite direction is defined as follows:

$$bpa_2(T^{A22}) = \frac{1 + cos(\phi)}{2}\chi^2_{22}, bpa_2(F^{A22}) = 0, bpa_2(\Omega) = 1 - bpa_2(T^{A22}) \tag{5.4}$$

(χ^1_{22}, χ^2_{22} are the scaling parameters for these *bpa*)

Another argument pro "threat" is based on the number of observed boats within a required towing distance from a vessel under consideration, and the type of the vessel defining the number of towing boats required. For a vessel type requiring two tug boats, the argument *Arg3* pro "threat" ("the number of towing boats is not the allowed one") is a conjunction of the following assumptions:

A31: Tug boat 1 is within the prescribed distance for a tug boat

A32: Tug boat 2 is within the prescribed distance for a tug boat

A33: Tug boat 3 is within the prescribed distance for a tug boat

A34: The vessel requires 2 tug boats

The belief in the argument is a combination of beliefs that there are 3 boats detected within the prescribed towing distance, and beliefs in the number of allowed boats, which comes from credibility of vessel ID recognition based on vessel characteristics. Beliefs that boats are within the towing distance is computed by an expression similar to Equation 5.2.

We also consider *Arg4* related to the rules of towing operations; specifically, "alignment of the towing boats is not correct". It is a conjunction of three assumptions based on the alignment of the boats. The beliefs on the assumptions are computed with expressions similar to Equations 5.3 and 5.4. *Arg5*, in turn, is a pro "threat" argument based on the fact that one of the boats is breaking a security perimeter. It is a conjunction of an assumption *A51* and a proposition *P5*:

A51: Boat is close to restricted access facility

P5: The facility perimeter must be protected

The *bpa* for *A51* is a function of the distance between the boat and the secured facility. To model de belief it is used an equation similar to Equation 5.2.

5.6.3 Results

This subsection shows simulation results on the scenario depicted in Figure 3. In the simulation, three tug boats (*s1, s2* and *s3*) seem to be towing a power-driven vessel (*s4*) of length 180m. The operation is carried out from the south of the GCN channel to the dock at the end of the SCW1 channel. Harbor rules state that *s4* only requires 2 tug boats, but in the simulation we have three candidates. *s1* and *s3* are not compliant to the harbor requirements in several stages of the trajectory, which makes it difficult to determine which one is a real tug boat. The most noticeable misbehaviors happen at the middle of the operation, where *s3* increases its speed over the limits allowed for the navigation channel, and at the end, where *s1* heads to the secured facility. Simulation data includes position, size, angle and

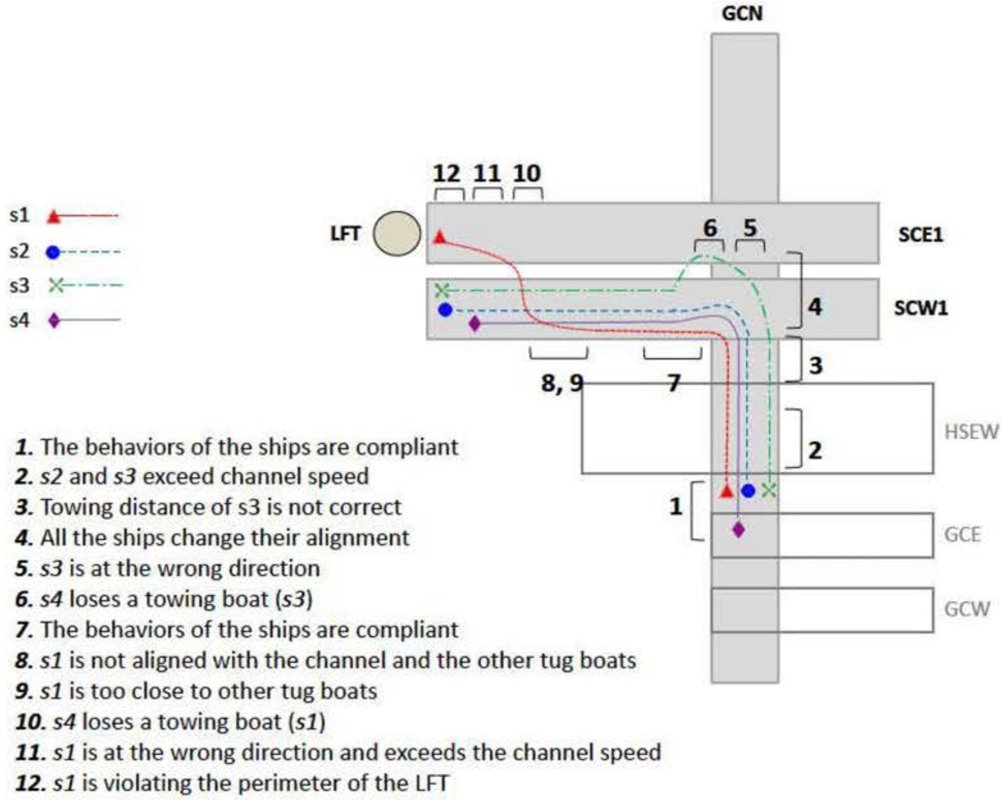


Figure 5.5: Simulated trajectories and behaviors in the harbor scenario

speed for each ship during 42 time instants (168 registers). Figure 5.5 shows in detail the ship trajectories and labels their behavior in order of appearance.

In Figure 5.6 we can see the tug boat *s3* increasing its speed at t=6 and exceeding the speed allowed for the channel. *s3* accelerates at t=[5, 9], and maintains a stable speed at t=[9, 13]. From t=14, the belief of the argument *Arg1* into hypothesis "threat" due to speed violation decreases, since the behavior is no longer incorrect. Similarly, *s1* accelerates at t=[36, 39] and then maintains a stable speed increasing the value of the belief. As expected, the actual belief values in these two cases are different, since the difference between the maximum allowed speed in the channel and the boat speed is larger in the latter.

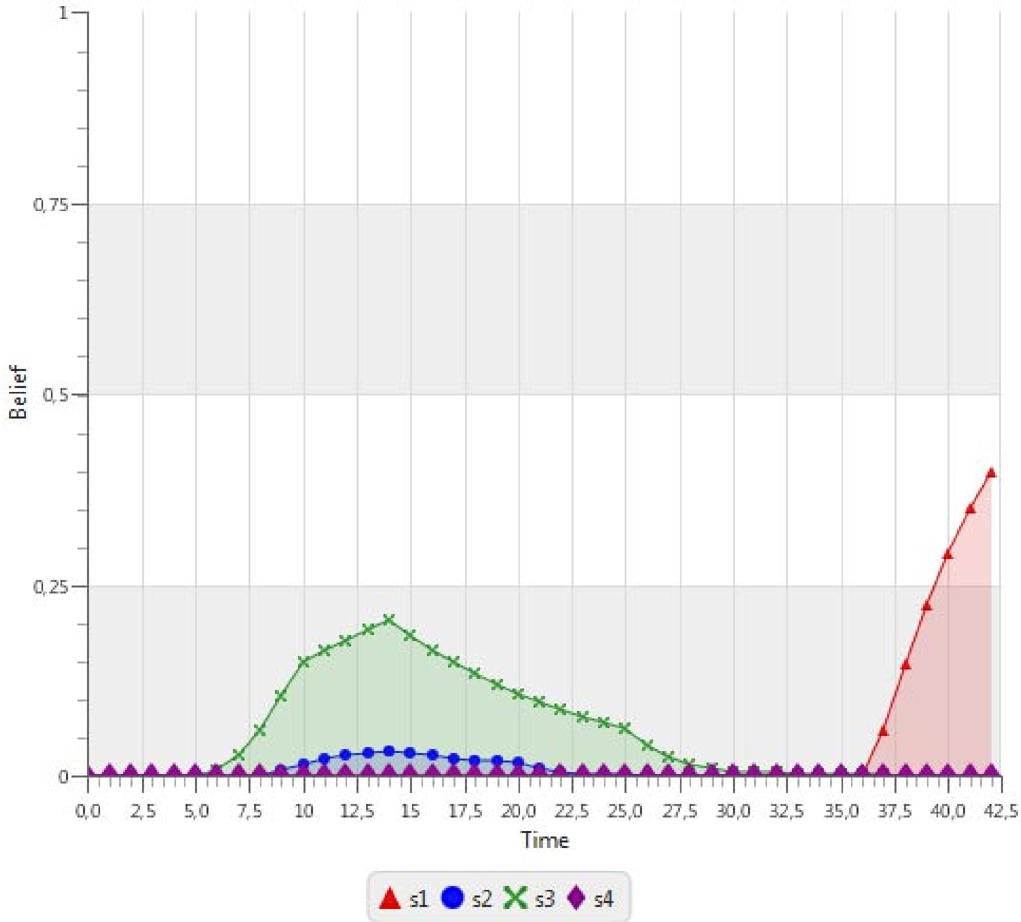


Figure 5.6: Dynamics of belief in *Arg1* (speed violations)

To bring the vessel to the dock, all the ships must turn left into the overlapping area of the GCN and SCW1 channels. During this maneuver, the ships are not aligned to the channels in which they are into. This causes an increment in the value of the beliefs into the argument pro "threat" hypothesis *Arg2*, corresponding to the violation of the navigation channel direction, at t=[13, 18], as depicted in Figure 5.7. Later, *s3* navigates against the direction of the SCE1 channel, which increases the value of the belief. It is also interesting to highlight that at the beginning of the simulation, *s4* is simultaneously inside GCN and HSEW. For some time, *s4* infringes the alignment with the HSEW channel, but since it is aligned with the GCN channel, the value of the belief does not increases.

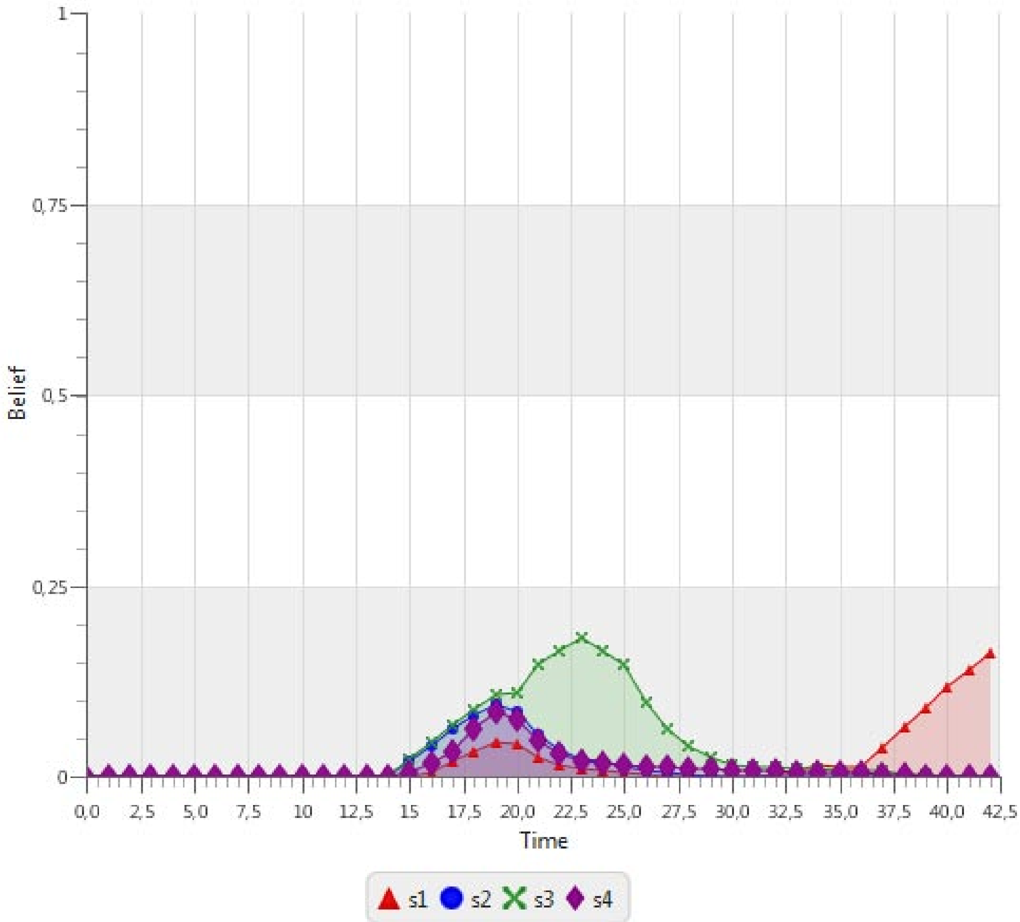


Figure 5.7: Dynamics of belief in *Arg2* (alignment to channel navigation directions)

Figure 5.8 shows the effects of the detected high speed of *s3* to the values of the beliefs in the arguments related to towing operations. As a result of *s3* acceleration, the distance between *s3* and the towed vessel *s4* increases starting at t=20. After a while, *s3* is not considered to be towing *s4*, because the distance exceeds the maximum value to which a boat can be involved in a towing operation. Therefore, the belief on the argument pro 'threat" *Arg3* decreases because the number of tug boats of *s4* is correct when *s3* is not considered a participant of the operation. A similar situation happens at the end of simulation, when *s1* moves towards the LFT.

Changes in the alignment of boats with respect to channels do not affect very much the belief in the argument *Arg4* related to alignment between towing boats, since boat trajectories in the simulation are consistent. As shown in Figure 5.9, the most noticeable situations are the trajectory deviation by *s3* at t=[17, 24] and *s1* at t=[31, 36]. Figure 5.10, in turn, depicts the dynamics of the belief in the argument pro "threat" hypothesis related to facility perimeter violations. As expected, *Arg5* quickly increases when *s1* approaches to the protected facility LFT.

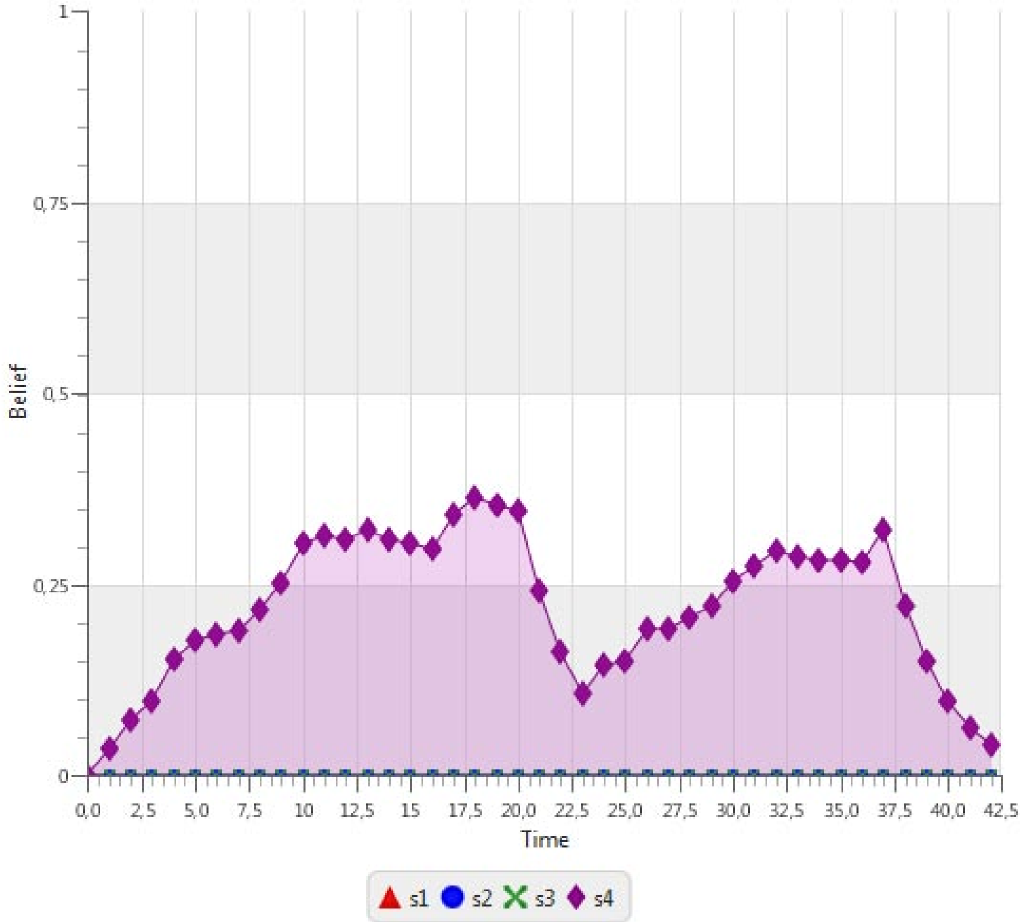


Figure 5.8: Dynamics of belief in *Arg3* (number of towing boats)

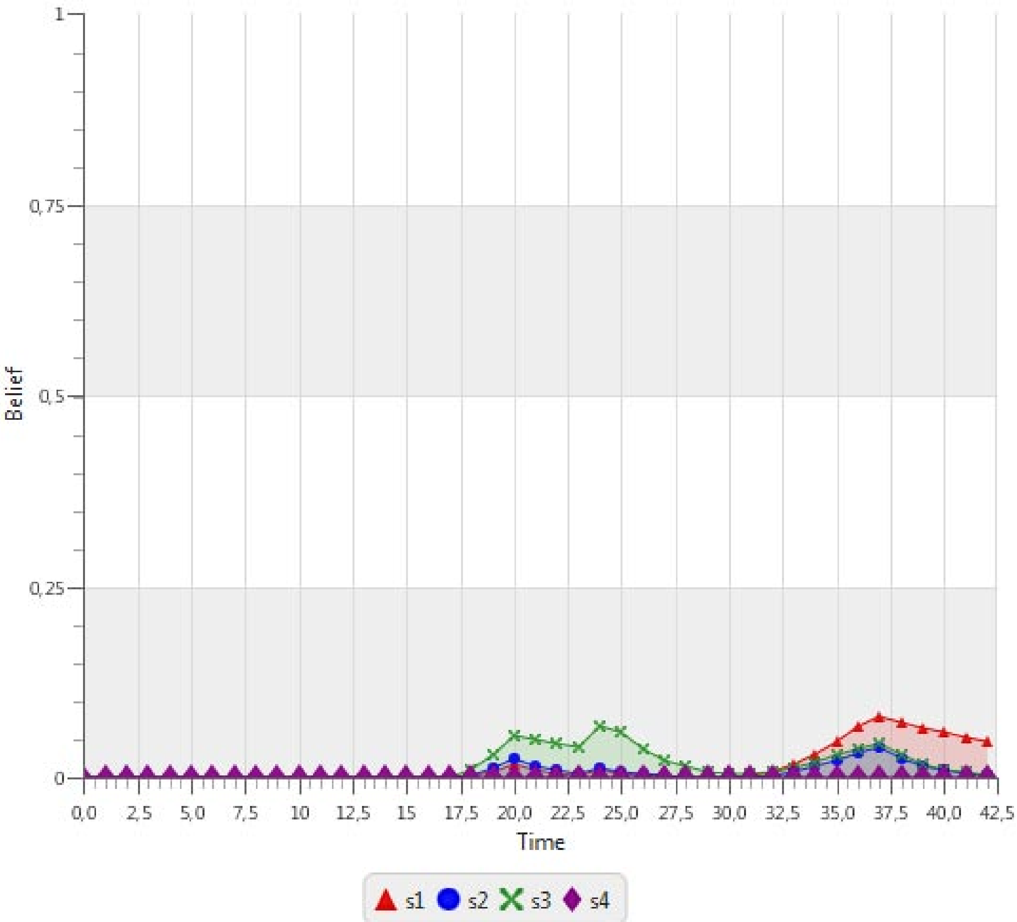


Figure 5.9: Dynamics of belief in *Arg4* (alignment between towing boats)

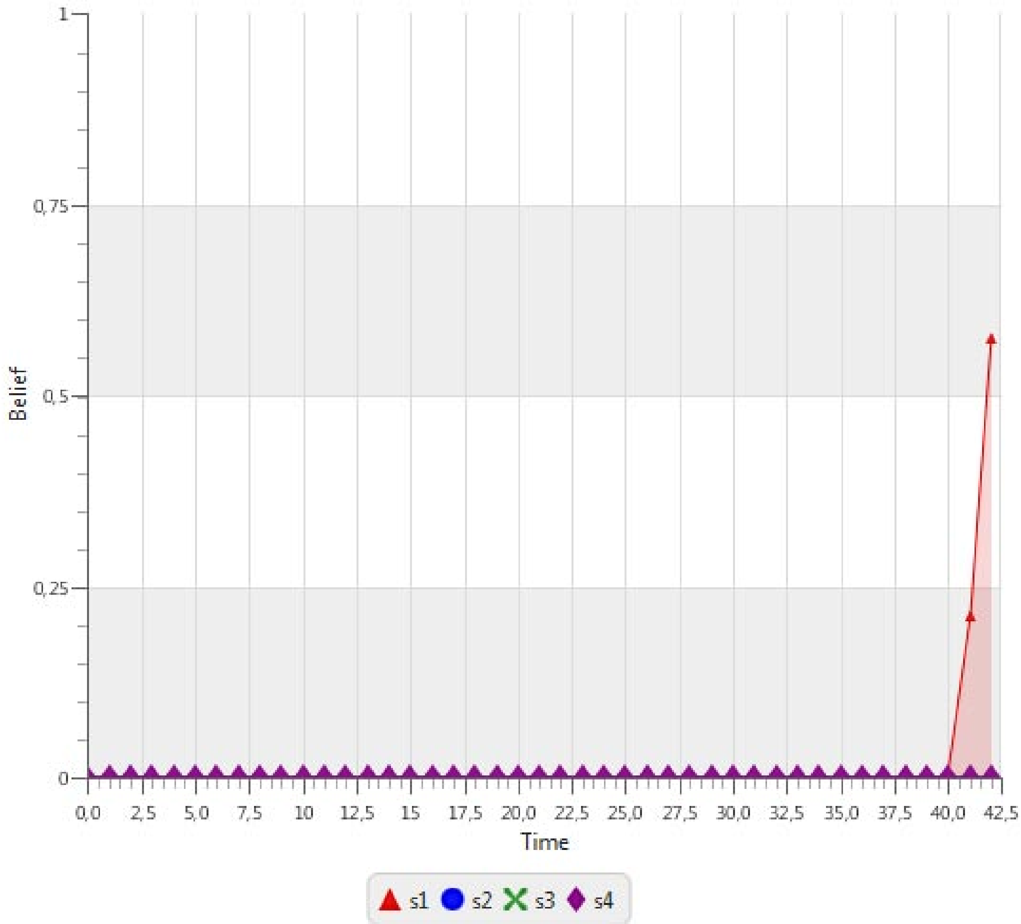


Figure 5.10: Dynamics of belief in *Arg5* (close to facility)

Figure 5.11 shows the overall belief into the hypothesis "threat". It depicts several situations of interest through the simulation. At t=16, *s4* exhibits a combination of non-compliant behaviors; namely, number of tug boats and alignment to channel navigation direction. *s3*, in turn, has an erratic behavior in t=[5, 24], including violations of speed, direction alignment and tow alignment violations. Nevertheless, the evidence accumulated in favor of the "threat" hypothesis does not reach enough relevance to be considered. In contrast, the threatening behavior of *s1* at t=[30, 42] results in triggering an alarm when it approaches the LFT area.

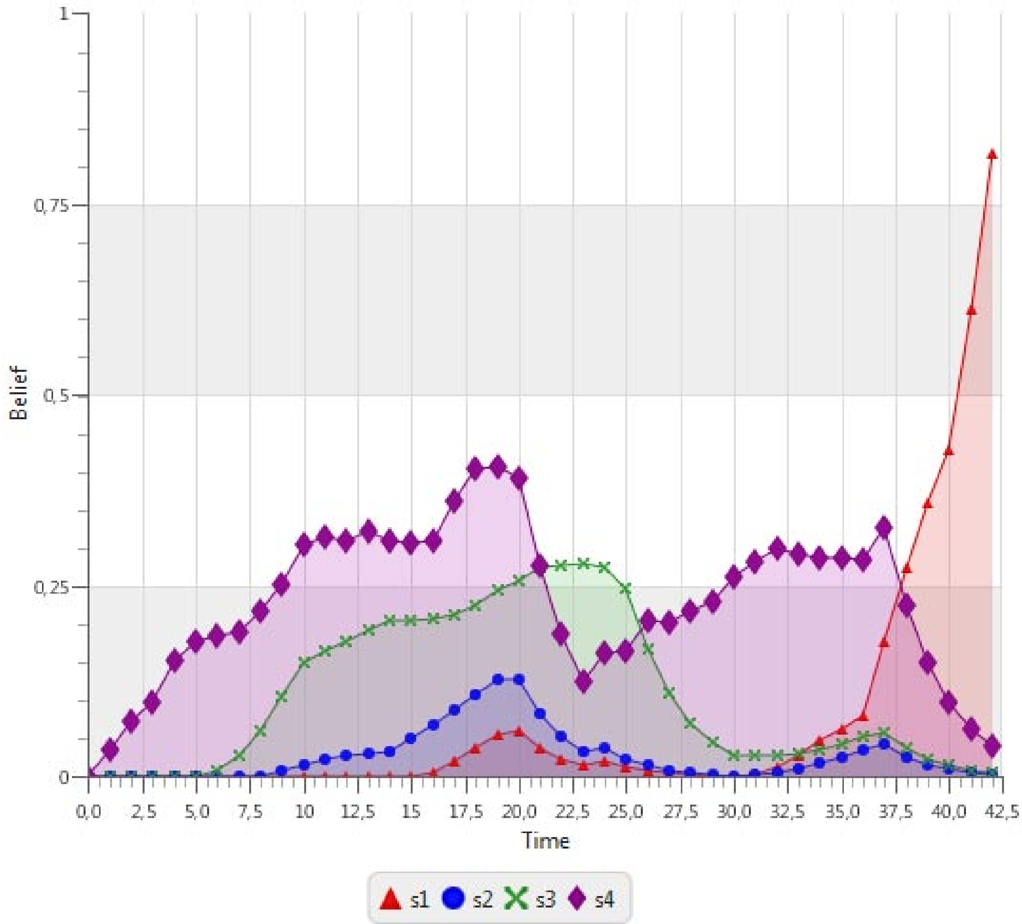


Figure 5.11: Dynamics of belief in "threat" hypothesis

6

Conclusions and future works

Some of the main tasks of a context-aware systems are to acquire, represent, and reason with general data to provide high level scene interpretations. This dissertation has described a framework that covers the required processes to symbolically represent and reason with context information and sensor data to achieve scene understanding as a first step towards the provision of customized functionalities. The leit motif of this thesis was to demonstrate the capability of this general framework to act as a basis on which develops novel applications in a wide range of areas imbibing in it cognitive layers when needed to extend its capabilities.

The cornerstone of the framework consist in an ontological model designed according to the JDL process model, the canonical specification to describe multi-sensor systems proposed by the Information Fusion research area. The framework can be included into the category of multi-level fusion systems, since it achieves a high-level abstract interpretation of the complete scene in terms of objects and situations using multiple datasources.

One of the main advantages of the proposed framework lies in the use of ontologies to represent and reason with the cognitive scene model. Ontologies support the creation of a model skeleton defining top-level concepts and relations, thus allowing domain-specific applications to extend and reuse it. In addition, the use of a common cognitive model facilitates the incorporation of new sensors in the case of working with a network of sensors, since they can communicate with a central node as long as they use the proper ontology to encode information. Information coming from contextual datasources can also be easily incorporated. In general, processing algorithms and techniques could be transparently replaced, which makes the framework more extensible. Besides, symbolic scene representations are more interpretable, which facilitates participation of human users in the system, as well as debugging and adjusting the algorithms.

As the formulation of models based on abstraction levels has led to the implementation of non-cohesive systems. This work has proposed new common, general and transverse knowledge layers among these levels with the goal of obtain a close interaction among semantically similar layers and the automatic generation of new knowledge. More specifically the proposal updates the cognitivist models towards qualitative spatial and mereological layers. These new layers had specific implementations as a dynamic topological approach, a theoretical taxonomy of mereological relations and the ties between them through data propagation patterns.

A key aspect of the framework is the different reasoning capabilities that is able to develop: Ontology-based reasoning to perform consistency checking and subsumption, engine support to develop deductive processes and abduct new knowledge using rules, spatial and part-whole reasoning for a more complete representation and understanding of the scene and BAS to deal with uncertainty situations. This work demonstrate that all these mechanisms can and should be orchestrated and each one should have a specific role when context-based systems are created.

According to the increasing reasoning capabilities the framework has been incrementally presented using applications corresponding to different subfields of context-aware systems; namely, AmI, SSP and surveillance systems.

The AmI prototype presents the framework from the persepective of visual sensor networks in a smart home application. The basic reasoning procedures performed by the framework consistency checking and rule base reasoning are applied for object identification, tracking enhancement and single/multiple camera scene identification. The graphical tool can be considered as a first step towards the incorporation of the human operator in the system –which is called Level 5 fusion. Application domain adaptations, such as the smarthome application specific ontologiy have shown that it is possible to take advantage of the general domain as the basis of a specific domain making their definition faster, in terms of developing time, and much more lighter.

To illustrate the functioning of the extended framework, a SSP application case study for live market research has been described, by presenting some examples of activity recognition representation and data propagation. These examples are able to represent semantically complex relationships through the interpretation of the users interactions with the context. In this regard, a general ontology-based model for formal representation of the human body is presented. The model has been embedded into the first prototype by relying on part-whole patterns and DOLCE recommendations. The proposal includes Kinect^{TM} skeletal view data representation with backward compatibility with the previous proposal. The main advantages of this model are the general representation for further domain extensions and the logical capabilities for automatic inference of high-level relationships. Both advantages provide support for more sophisticated activity analysis.

The application of the framework to the harbor surveillance example depicts a complete situation assessment. Contextual information is included into this model in the form of definitional classes, which are used to classify entities, and deductive rules used to infer discrepancies with respect to the normal operations. Rules encode the restrictions defined by the port authorities based on the geometrical configuration of the harbor and navigational channels. The system utilizes an uncertainty reasoning method based on the Belief Argumentation System to identify the deviations from the normalcy model that truly correspond to threatening situations and avoid false alarms due to spurious errors.

Despite all these advantages the proposal presents some limitations. The main one is that we have shown prototype implementations of the system with simplified rules, but real-world applications must be still developed and tested. In addition, we have overlooked some problems that appear in a real application; e.g., errors in the tracking procedure, latency produced by the reasoning procedures, overhead due to irrelevant minor changes between scenes and so forth.

Scalability issues of the topological reasoning module must be addressed as well. Applying a context selection technique, can optimize context exploitation by allowing the system to only search in the subset of the context model that is relevant to the current situation.

Building a fully-deployable implementation of the system requires solving several additional problems that are outside the scope of this thesis. It is necessary to develop a better integration of higher-level fusion processes (classification of abnormal situations, abductive reasoning) with lower level tasks (object detection, tracking and identification). This requires designing a detailed model of inter-level procedures, including processes for quality control and multiple feedbacks to improve the global performance of the system.

Another interesting research direction is the incorporation of additional uncertain and vague representation reasoning formalisms. Classical ontologies do not provide support for this kind of knowledge, which is inherent to applications involving abductive reasoning procedures: sensor data may be imprecise; local scene interpretation procedures may be uncertain; information fusion might be partially trusted; etc. Furthermore, it may be interesting to add imprecise knowledge to the cognitive model; e.g., imprecise spatial predicates (RCC predicates that hold to a certain degree), additional fuzzy entities (imprecise definitions of entities) and spatio-temporal relations (close, far, recently, etc.).

Unfortunately, a detailed comparison of the overall proposal is not possible at this moment, because public implementations, datasets, scenarios and criteria for the evaluation of higher-level fusion systems are scarce, if not inexistent. The creation of such evaluation framework, a task addressed by the ETURWG group, is a prospective direction for future research.

A

Context in context-aware systems and Information Fusion

The Webster dictionary defines context as "the interrelated conditions in which something exists or occurs" or "the parts of a discourse that surround a word or passage and can throw light on its meaning"[1]. The concept of context has been studied in many research fields (see for example (Bradley and Dunlop, 2005)).

One of the first approximations to the formalization of the notion of context in Artificial Intelligence is due to McCarthy (McCarthy, 1993), who proposed the use of the relation ist(c, p) to represent that a given proposition p is true in the context c. Sowa extended this theory with the dscr(x, p) relation (Sowa, 1995), which states that p describes entity x. Since x can be a situation, dscr semantics subsume those of ist. Giunchiglia defines a similar epistemological framework in which a context is a subset of the complete state of an entity that is used in reasoning to solve a task (Giunchiglia, 1993). It has been proved that these multi-context logics are more general than ist-based formalisms (Serafini and Bouquet, 2004).

The first conceptions of context-aware systems took place at the beginning of the 90s in several non-related researches (Harter and Hopper, 1994), (Schilit et al., 1993), (Spreitzer and Theimer, 1993), (Want et al., 1992), (Weiser, 1991), since then, many definitions have emerged trying to give a global view covering the increasing aspects of context. First definition was introduced by (Schilit and Theimer, 1994) in 1994 as *"the ability of a mobile user's applications to discover and react to changes in the environment they are situated in"*. This definition was founded on a concept of context which only took into account the identity of the people, located-objects and services. However, along the years, such enumeration has been changing becoming more general and accurate. In 1997 Ryan et al. (Ryan et al., 1997) described context as the computer's environment, such as location, time, temperature or user identity. Afterward, this definition was updated by the notion of user's context including but not limiting the concept to *"emotional state, focus of attention, location and orientation, date and time of day, objects and people in the user's environment"* (Dey, 1998). Unfortunately, these kind of definitions are only subsets of features associated to context and thus they will never cover future aspects of the concept. Other ways to define context can be found

[1]Webster dictionary. http://www.merriam-webster.com/dictionary/context Last accessed December 2015

throughout the literature, for example, Schilit et al. (Schilit et al., 1994) considered three aspects of context: where you are, who you are with, and what resources are nearby and introduced the idea of non-stationary context. Pascoe (Pascoe, 1998) estimated that context is the subset of physical and conceptual states of interest to a particular entity. Although these definitions are very general, they are still incomplete.

The current starting point to define context is the statement given by Dey et al. (Dey and Abowd, 2000) *"Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."* From this general conception the classification of the different types of context is more affordable. For instance, Gustavsen (Gustavsen, 2002) considered a non-closed list of categories of context including the user-context, system-context, application-context, social-context and historicalcontext. Kappel et al. (Kappel et al., 2002) separated the physical context low level layer which constantly update the environment state through sensors and the logical context –abstract data which enrich the semantics of physical context information. This research also identified three types of physical context: (i) natural context, location and time data; (ii) technical context, device, network and application data; (iii) social context, the knowledge about the user.

During the last years new challenges have been established based on the refinements did in the context conception. The key aspect of these refinements is that, from an ubiquitous perspective, context is part of a process. Context-aware systems should be able to interact with different agents during this process and adapts its behavior to changing situations by means of cooperation and reinterpretation rather than sophistication (Coutaz et al., 2005). This process has to deal with four issues: (i) collect useful and enough information for its purposes; (ii) choose the smallest set of relevant factors for further analysis; (iii) generate a response understanding the latent relationships in the data; (iv) evaluate the effectiveness of the results (Yujie and Licai, 2010).

In the Information Fusion community, context has been considered from different points of view. One of them, which seem to be prevalent, is to refer to external knowledge that is useful or influences the fusion processes, including background knowledge (e.g. tactics, doctrine), situation-specific knowledge (e.g. terrain), existing reports and databases, and so forth (Powell et al., 2006), (Kandefer and Shapiro, 2008). Sycara et al. state that part of the context are the significant features or the history of a situation that influence the features of other situation, as well as the expectations on what is to be observed and the interpretation of what has been observed (Sycara et al., 2009). They also propose the HiLIFE (High-Level Information Fusion Environment) fusion model for battlefield management. To these authors, situational context is a "first class entity", but not exactly in the sense of McCarthy. In their sense, it is rather a computable description of the terrain elements, the external resources and the possible inferences that is essential to support the fusion process.

In (Steinberg and Rogova, 2008), (Strang and Linnhoff-Popien, 2004), several major types of context models were considered, of which the three ones most applicable to data fusion can be characterized as key-value models, ontology based models and logic-based models. Key-value models are the simplest way of representing context. They provide values of context attributes as environmental information and utilize exact matching algorithms on these attributes. These models may suffice for use in Level 1 fusion to work with data constraints (George et al., 2009), but they lack capabilities for complex context representation required by higher-level fusion. Ontology based models provide a formal way for specifying core concepts, sub-concepts, facts and their inter-relationships to enable realistic representation of contextual knowledge (Nowak, 2003), (Little and Rogova, 2009), (Kokar et al., 2009). Current approaches to ontology-based context modeling can be classified into three main areas: contextualization of ontologies, ontology design patterns, and context-aware systems (Gómez-Romero et al., 2011c). Ad hoc logic-based models can be applied to extend or replace ontologies in knowledge-intensive applications. They represent context as facts and information inferred from rules. These models are generally more expressive, and allow for the development of more sophisticated representations and reasoning procedures.

A.1 The evolution of context representation formalism in context-aware systems

The desired features for context-aware systems are closely intertwined with the evolution of context modeling techniques. These techniques should meet different requirements inherent to the nature of these applications. For example, a well designed context information model must be able to adapt their behavior to give appropriate response in a variety of scenarios, has to cope with several types of context, different updating frequencies, semantically leveled data, and so forth.

In the beginning, context models were based in key-value pairs defined over markup languages such as XML. More complex languages based on graphical models replaced key-value approaches, such as the Unified Modeling Language (UML); e.g. (Sheng and Benatallah, 2005). The most prominent example was the Context Modeling Language (CML) (Henricksen et al., 2004) whose origins were rooted with the Object Role Modeling (ORM). The CML provided some modeling constructs to represent different kinds of context facts, non-perfect data, dependencies between context fact types and historical data. CML worked with three-valued logic and a grammar for formulation situations (Henricksen and Indulska, 2006). However CML did not allow to represent the context types using a hierarchical structure, nor gave support for interoperability between models (Bettini et al., 2010).

Due to these limitations other alternatives were explored ranging from object-oriented models to logic-based models (Baldauf et al., 2007). The formers, tilted towards knowledge representation, had to provide communication interfaces for inference systems (Hofer et al., 2002). The latters, more concerned about the reasoning procedure, had to declare context as facts and define inference rules (McCarthy and Buvac, 1997).

In those years a new language based on XML, called Resource Description Framework (RDF)[2] was able to represent, relate and constraint context data through hierarchical structures. These models formally limited and unable of using general purpose reasoning techniques evolved towards ontologies. Ontologies presented an optimum trade-off between representation and reasoning capabilities, adding support for consistency checking and subsumption reasoning mechanisms –given two concepts which one is more general– to its inherent features of expressiveness and interoperability. Since then, many context-aware

[2] Resource Description Framework (RDF). http://www.w3.org/RDF/ Last accessed December 2015

approaches opted to represent context data coming from sensors, inference processes and users in structured description logic (DL) expressions. The DAML+OIL ontology language (Horrocks, 2002) was used at the beginning to carry out the first context-aware approaches based on ontologies. A good example of the DAML+OIL capabilities is the GAIA middleware (Ranganathan et al., 2003), which was able to derive new context data by means of rules and statistical learning. DAML+OIL was substituted by the first version of the Ontology Web Language (OWL)[3]. The most prominent ontologies in this language were SOUPA (Chen et al., 2004b) and CONON (Zhang et al., 2005) designed for pervasive and smart home environments respectively. These ontologies were used later in several architectures for context awareness, such as, the COntext Broker Architecture (CoBrA) (Chen et al., 2004a) and the SOCAM (Gu et al., 2004b) middleware. The OWL's second version[4] is succeeding their predecessor as the building-block of new context-aware systems. OWL 2 overcome some important expressiveness limitations of OWL 1 in terms of relationships and rule-based reasoning. The research by Riboni et al. (Riboni and Bettini, 2011) is an illustrative example of generational renewal of context-aware approaches towards OWL 2.

[3]OWL 2 Web Ontology Language Document Overview. http://www.w3.org/TR/owl2-overview/ Last accessed December 2015

[4]OWL Web Ontology Language Document Overview.http://www.w3.org/TR/owl-overview/ Last accessed December 2015

B

Qualitative spatial representations

Qualitative spatial representations (QSR) have become a key component to represent and reason with spatial knowledge because of its proximity to the way humans define the spatial knowledge. A qualitative spatial knowledge model uses a formal vocabulary to describe the relations between the entities of the domain in a specific aspect of the space. For instance, abstract representations of spatial and topological properties –'A is inside B' or 'A is above B'– are close to the natural language, and can be exploited to bridge the semantic gap between symbolic and numerical representations.

There are several works in the literature that study the cognitive aspects of the space; e.g. topology, direction and distance, as well as, formal theories that focus on the representation of their semantics and the properties of these reasoning procedures; e.g. decidability. Topological approaches aim to qualitatively describe the spatial relations between subsets (or regions) of a topological space. The first formalizations are due to Whitehead (Whitehead, 1929) and Clarke (Clark, 1985; Clarke, 1981). These approaches are based on the extension of the basic connection relationship by applying logical theories to obtain additional well-defined relations.

B.1 Region Connection Calculus in ontologies

The Region Connection Calculus (RCC) is one of this axiomatizations in first order logic (Randell et al., 1992), (Renz, 2002). The basic RCC theory assumes just one primitive dyadic relation *C(x, y)* –read as 'x connects with y'–, where *x* and *y* denote spatial regions. This relation is reflexive and symmetric. Many different subsets of relations can be defined by using the RCC theory. The most popular is a set of eight base relations called RCC-8, since it can be encoded in propositional modal logic (Bennett, 1994), and therefore it is decidable. An alternative approach is the 9-intersection (Egenhofer, 1991), which defines nine binary relations including exterior, interior and boundary relations between regions. Unfortunately, the 9-intersection has not been proved to be decidable.

The most used version of RCC is RCC-8, which defines eight relations: DC (is disconnected from), EC (is externally connected with), PO (partially overlaps), TPP (is a tangential proper part of), NTPP (is a non-tangential proper part of), TPPi (inverse of TPP), NTPPi (inverse of NTPP) and EQUAL. These relations have been proved to compose a jointly exhaustive and pairwise disjoint set. Similar sets of one, two, three, and five relations are also defined (respectively, RCC-1, RCC-2, RCC-3, and RCC-5).

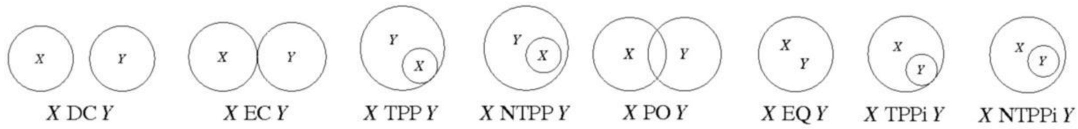


Figure B.1: RCC-8 relations

Not surprisingly, these factors have favored the use of RCC-8 in ontology-based approaches. First attempt was from Katz and Grau (Katz and Grau, 2005a), who carried out a translation from the feature relations of RCC-8 to the OWL language. The main problem was the absence of reflexive roles in OWL, which is one of the key assumptions of the RCC relations. According to the authors, the problem could be easily solved by using an extension of the description logic language. However, this kind of approach has additional problems, as described in (Stocker and Sirin, 2009): a huge amount of TBox axioms are generated as a result of the definition of the RCC-8 roles and the axioms specifying the non-emptiness of some regions.

Next version of the language, OWL 2 –based on the Description Logic $\mathcal{SROIQ}$ (Horrocks et al., 2006)–, include reflexive roles. In (Grutter and Bauer-Messmer, 2007a), (Grutter and Bauer-Messmer, 2007b), a translation of the RCC-8 into OWL 2 is presented. This approach addressed new problems. For example, OWL 2 does not allow the definition of a concept as an individual, and therefore regions have to be represented as individuals. As a result, the spatial domain cannot be represented as a strict set of concepts and relations. Another issue is that OWL 2 does not support all the role inclusion axioms used in the composition tables needed for the RCC reasoning. According to (Hogenboom et al., 2010), RCC-8 also requires role negation, conjunction, and disjunction, as well as complex role inclusion axioms. Using $\mathcal{SROIQB}_s$ logic (Rudolph et al., 2008), which adds role boolean operators to $\mathcal{SROIQ}$, some of these needs can be covered. Unfortunately, this logic does not support complex role inclusion axioms on the right hand side nor boolean role operators on complex roles.

Other proposals have faced the problem at a knowledge representation level, instead of at a formalism level. Specific components named RCCBoxes have been defined to manage spatial relationships. These RCCBoxes have predefined RCC relationships and composition tables, and use OWL 2 –they need support for negation roles to define a disconnected relation if none of the other relations are detected. RCCBoxes have been implemented in the Pellet (Stocker and Sirin, 2009) reasoner. In addition the RACER (Renamed Abox and Concept Expression Reasoner) reasoner[1] have contributed to the QSR representation by means of static substrates, unfortunately this solution do not have reasoning capabilities.

Most practical ontology-based approaches that require spatial properties are geographical information systems (GIS). These systems have a wide variety of applications; e.g. disaster management (Klien et al., 2006), data retrieval (Wiegand and García, 2007). Nevertheless,

[1]RACER engine web page. http://www.ifis.uni-luebeck.de/~moeller/racer/ Last accessed December 2015

most of them are more focused on representation issues, since they use ontologies to improve interoperability between heterogeneous systems. In the last years, new systems to query over spatial objects and features have appeared, though their expressiveness is quite limited; this is the case of AllegroGraph[2] and Geospatialweb[3]. A more general approach is (Van Hage et al., 2010), which offers a tight integration with OWL reasoning procedures and implements geometric operations supported by external libraries.

[2] Geospatial Tutorial for AllegroGraph. http://www.franz.com/agraph/support/documentation/current/geospatial-tutorial.html Last accessed December 2015

[3] GeoSpatialWeb. http://code.google.com/p/geospatialweb/ Last accessed December 2015

C

Ontologies under uncertainty conditions

Perception of reality has an unambiguous tendency towards incompleteness and vagueness in its observations. Devices, humans and data transformation processes add imprecision and errors to measurements taken in the real world. These features associated to data arise when we try to represent information with a certain accuracy. They are the causes of mistakes, malfunctions and even contradictions in systems' reasoning.

For the above reasons, different entities in a context-based system must be able to reason about uncertainty. However, traditionally, the reasoning associated to description logic and ontologies have been limited by a combination of two factors: i) the lack of suitable inference mechanisms able to manage imperfect knowledge; ii) the open world assumption. As an alternative to these limitations, systems founded on these technologies developed hybrid approaches in which logical or ontological knowledge representation and rule-based reasoning were executed separately. Some good examples are CoBrA (Chen et al., 2004a) with a loose integration with Horn clauses rule-based and ontological reasoning are executed separately and 2*3CM (Yu et al., 2008) tightly integrated with the Semantic Web Rule Language (SWRL) [1]. Unfortunately, even though the rule systems in loose integrations can derive new knowledge, this knowledge never return to the ontological reasoner. As a result, the rule-based reasoning cannot be exploited to derive other implicit information. In addition, the assertion of new knowledge coming from rules, either in tight and loose approaches, causes consistency and decidability problems.

Uncertainty reasoning in context-based systems has two primary objectives: (i) improve the accuracy of the context data given; (ii) infer additional context data from higher levels. There are several reasoning techniques to manage uncertainty such as, fuzzy logic (Zadeh, 1999), Bayesian networks (Pearl, 1988a) and Dempster-Shafer theory (Shafer, 1976). These techniques, have been mainly applied to specific domain problems in context-based reasoning to carry out improvements in the accuracy level of sensed data. Since our aim is inferring data from higher levels, we are interested in approaches which combine ontologies and reasoning with techniques able to manage uncertain context information (Bettini et al., 2010).

[1] SWRL: A Semantic Web Rule Language Combining OWL and RuleML. http://www.w3.org/Submission/SWRL/ Last accessed December 2015

The aim of the W3C Uncertainty Reasoning for the World Wide Web Incubator Group (URW3-XG) is a full realization of the World Wide Web as a source of processable data and services demands formalisms capable of representing and reasoning under uncertainty. For this purpose the URW3-XG published a final report[2] distinguishing the current techniques and guidelines to incorporate ambiguity, inconsistency, randomness and vagueness to web ontologies. The following subsections analyze the current alternative approaches combining ontology representation and uncertainty reasoning that may be incorporated to Information Fusion systems.

C.1 Bayesian Networks

Bayesian networks are a powerful graphical language for representing probabilistic relationships among large numbers of uncertain hypotheses. Bayesian networks assume a simple attributevalue representation in which each problem involves reasoning about a fixed number of attributes. This theory considers that all hypotheses are exclusive and exhaustive which is not frequently true in context-based systems. In addition this reasoning technique cannot be used when the number of random variables changes along the time. This is a major drawback in environments where the number of entities and relationships is not accurate and even the own definition of these entities and relationships may be uncertain.

First works were focused in allowing a probabilistic definition of concepts and roles in description logics. Unfortunately these approaches did not provide support for assertional knowledge. Some examples are (Heinsohn, 1994), (Koller et al., 1997) which represented probabilistic component using a Bayesian network for each class and (Yelland, 2000) which combines a restricted description logic with Bayesian networks.

The earliest relevant attempts to embed Bayesian reasoning in ontologies were BayesOWL (Ding and Peng, 2004) and the work by Gu et al. (Gu et al., 2004a). Both tried to extend ontologies allowing probability annotations and translating them through rules into the directed acyclic graph (DAG) of a BN. Unfortunately, standard Bayesian networks lack the expressive power to fully represent ontologies. This fact limited the use of these approaches to specific problems. Another interesting work was the DL reasoner Pronto which was able to represent and reason about uncertainty in both, generic background knowledge and individual facts. The major drawback of this tool was the scalability.

Current research in this field is mainly focused in probabilistic ontologies. A probabilistic ontology is a formal knowledge representation which includes among other features: i) statistical regularities that characterize the domain; ii) inconclusive, ambiguous, incomplete, unreliable, and dissonant knowledge related to entities of the domain; iii) uncertainty about all the above forms of knowledge. Some examples include the upper ontology PR-OWL (Costa, 2005) and KEEPER (Pool and Aikin, 1994). The former based on Multi-entity Bayesian networks (MEBN) –a formal system that integrates First Order Logic (FOL) with Bayesian probability theory– is one of the few proposals that has an open source project associatedXSB[3] and includes reasoning implementation packages (Costa et al., 2008).

[2]Uncertainty Reasoning for the World Wide Web Final Report. http://www.w3.org/2005/Incubator/urw3/XGR-urw3-20080331/ Last accessed December 2015

[3]UnBBayes. http://sourceforge.net/projects/unbbayes/ Last accessed December 2015

C.2 Fuzzy Logic

Fuzzy logic allows for representing and processing data with a certain degree of truth about an imprecise piece of information. While in uncertainty there is an associated probability for each possible world, –the probability quantifies the plausibility that a world is true or holds with the actual– in fuzzy logic we deal with vague statements and try to measure to what extent a statement is into a truth value space between false and true. Some functions generalize the ordinary logical operators namely, conjunction, disjunction, implication, and negation. To define satisfiability notions in fuzzy knowledge bases, statements and degrees of truth are related with inequalities: at least ($\geqslant$), at most ($\leqslant$), greater than ($>$) and lower than ($<$). Fuzzy description logics restrict the truth values of concept assertions, role assertions, concept inclusions, and role inclusions that are true with some degree in the closed interval [0,1].

The first remarkable work in this field was due to Yen (Yen, 1991), who generalized the subsumption mechanism over terminological knowledge. However, a very restricted version of the $\mathcal{ALC}$ sublanguage is used and no fuzzy assertional knowledge is addressed in this work. Similar drawbacks were noted in (Tresp and Molitor, 1998), nevertheless this work presented a more general fuzzy extension of $\mathcal{ALC}$ and implemented a tableaux algorithm to compute subsumption degrees. Hölldobler et al. (Hölldobler et al., 2002) developed a complete reasoning algorithm for the subsumption problem extending the fuzzy $\mathcal{ALC}$ (Straccia, 1998), (Straccia, 2001) with fuzzy modifiers.

More expressive description logics were extended with fuzzy formalisms. Good examples are: the $\mathcal{ALCQ}$ description logic where fuzzy quantifiers were also introduced (Sánchez and Tettamanzi, 2005), (Sánchez and Tettamanzi, 2006); the $\mathcal{SHOIN}(\mathcal{D})$ language –the logic behind OWL DL– whose semantics for a fuzzy extension were first introduced (Straccia, 2005) and then fully defined (Stoilos et al., 2005).

Some advances towards reasoning in fuzzy $\mathcal{SHOIN}(\mathcal{D})$ were carried out in the last years. Tableaux calculus algorithms have been developed for fuzzy $\mathcal{SHIN}$ (Stoilos et al., 2007) and fuzzy $\mathcal{SHI}$ with fuzzy general concept inclusions (Li et al., 2006). Current reasoning proposals with implementation are: the tandem fuzzyDL[4] and Fuzzy OWL 2[5], the former includes a reasoner for fuzzy $\mathcal{SHIF}$, the latter is a Proteǵeṕlug-in to build fuzzy ontologies in OWL2; FiRE[6] is a fuzzy reasoner based on an extension of DL $\mathcal{SHIN}$ with fuzzy set theory; and ONTOSEARCH2 (Pan et al., 2008) a scalable query engine for fuzzy DL-Lite ontologies.

C.3 Dempster-Shafer

DempsterUShafer theory, also known as Evidence theory is often presented as a generalization of the probability theory. This theory allows the representation of evidence from multiple

[4]The fuzzy DL system. http://gaia.isti.cnr.it/~straccia/software/fuzzyDL/fuzzyDL.html Last accessed December 2015

[5]Fuzzy Ontology Representation using OWL 2. http://gaia.isti.cnr.it/~straccia/software/FuzzyOWL/index.html Last accessed December 2015

[6]Fuzzy Reasoning Engine FiRE. http://image.ntua.gr/~nsimou/FiRE/ Last accessed December 2015

sources and the combination of such evidences to calculate the belief for a given hypothesis. Major advantages of the Dempster-Shafer theory are: i) explicit representation of ignorance; ii) the possibility of use of subjective belief instead of a priori probabilities in case they are not available; iii) belief masses can be given either to single and sets of hypothesis; iv) there are additional belief measures (e.g. belief and plausibility) which give different points of view about the support of the data to each hypothesis.

The formal theory is based on a frame of discernment which houses all the hypotheses under consideration. These hypothesis are mutually exclusive and exhaustive. The power set is the set of all possible sub-sets of the frame of discernment, including the empty set. Different sources can assign a belief masses to each sub-set in the power set. These masses can be combined by using Dempster's Rule of Combination in order to update the belief of the sub-sets. The complexity of the combination rule is one of the major drawbacks of these theory.

Current alternatives combining Dempster-Shaffer theory and ontologies are scarce. The high computational complexity mentioned above and the slow classification process of ontology reasoners make this combination unattractive. However, simplification algorithms (Barnett, 2008; Gordon and Shortliffe, 1985) and other combination rules (Smets, 1990; Yager, 1987) have been proposed that try to reduce the overhead associated with the computational complexity.

Despite the fact that the combination of ontologies and belief theories has not been widely studied, there are some early works that try to take advantage of the strengths in both fields. Hoist et al. (Hois, 2007) advocate for a combination of OWL ontologies and the Dempster-Shafer theory separating the strict and well-defined background knowledge and imprecise knowledge of a situation in the TBox and the ABox respectively. A concrete proposal can be found in (Schill et al., 2009) where the authors develop a system to discern between different rooms with a certain degree of belief using a visual recognition system, an ontology framework and the opinion of some experts. Another relevant work is (Zhang et al., 2009), the proposal includes a context-aware architecture and a strategy which extend Dempster-Shafer theory for context reasoning ($DSCR^x$) which considerably reduces the computation overhead, but not the combination rules complexity, by selecting evidence with the highest beliefs relying on the most related contexts. A theoretical work for reasoning with ontology constructors is BeliefOWL (Essaid et al., 2012). This research extends the OWL ontology classes with belief masses and applies a translation to get a directed evidential network (DEVN) (Yaghlane et al., 2003)–a graphical model to represent belief functions independencies. The major drawbacks of this work are that only consider including uncertainty in classes and there are not available tools to test it. As an alternative Bellenger et al. presented an approach similar to PR-OWL (Bellenger et al., 2012). Starting from an upper ontology called DS-Ontology, one can build domain-specific uncertain concepts whose individuals are linked with individuals representing a frame of discernment. In contrast to other studies uncertain instances can have both uncertain classes and properties. Belief masses and information sources are related to the corresponding elements of the power set of candidate instances. To enable classical evidential combination and decision processes, and due to the fact that candidate instances are not necessary disjoint from each other, some semantic inclusion and disjointness operators. These operators, based on similarity functions, are defined in order to map the universal set of candidate instances with a frame of discernment in the Evidential theory.

C.4 Belief Argumentation Systems

Let $\Theta = \{\theta_1^t, \dots, \theta_K^t\}$ be the set of hypotheses that are considered at the time instant t. Since we assume that anomaly can be the result of the insufficient quality (reliability, uncertainty) of the information on vessels and their behavior, this set of hypotheses includes a hypothesis representing "normal operations". The set of hypotheses $\Theta = \{\theta_1^t, \dots, \theta_K^t\}$ may not be exhaustive, since not all the causes of anomaly may be included in the frame of discernment and some of them can be unknown or even unimaginable (open world assumption). This means that plausibility of an unknown hypothesis can be different to zero.

There are two major types of models of reasoning under uncertainty: graphical models, such as Bayesian and causal networks (see, e.g. (Nicholson and Brady, 1994), (Pearl, 1988b)), and logic-based models. Since normal situation is based on context of normal operations expressed in rules, we select here a logic-based model. One of the logic-based paradigms that can be considered for abductive reasoning under uncertainty is the Belief-based Argumentation System (BAS)[7], a generalization of the Probabilistic Argumentation System (PAS). Following (Haenni et al., 2000), we describe PAS as a hybrid approach that combines logic and probability theory. It aims at assessing hypothesis about present or future worlds by relying on available uncertain, unreliable, incomplete and contradictory knowledge. Logic represents the qualitative part of PAS. It is applied to determine arguments that support (i.e., *in favor*) and refute (i.e., *against*) each hypothesis. An argument is a conjunction of propositions and uncertain assumptions coupled with *a priori* probabilities of their trueness that make a hypothesis true or false. The probabilities that the arguments are valid are combined to obtain the quantitative judgment on the validity of the hypothesis, which is then used to decide whether it can be accepted, rejected, or knowledge is not available to make an appropriate judgment at this time.

Precise knowledge of *a priori* probabilities for assumptions is hardly available in the uncertain dynamic maritime environment, in which different and even unimaginable behaviors (types of threat) can occur. Therefore, they have to be replaced by dynamic subjective beliefs. Moreover, $P(A)$ –additive subjective belief that assumption A is true based on expert subjective opinion– is not generally 1-$P(\neg A)$, because of this high uncertainty. Consequently, PAS needs to employ sub-additive subjective belief measures of the form $Bel(A)+Bel(\neg A)\leq 1$. This sub-additive property makes it possible to explicitly express ignorance, and does not force one to reduce total uncertainty to the assumption that all the hypotheses under consideration are equally probable. Thus the belief theories allow for representing only our actual knowledge "without being forced to overcommit when we are ignorant" (Barnett, 2008).

In addition, the open world assumption, in which $Bel(\emptyset)$ may not be equal to zero, also requires an uncertainty representation allowing a non-exhaustive set of hypothesis, which calls for the Transferable Belief Theory (Smets, 1990) as an uncertainty framework in BAS. The dynamic beliefs assigned to the assumptions are based on current context and observations. The beliefs are approximated by a function of the estimated values of attributes

[7] The formal description of the Belief-based Argumentation System below follows the explanation introduced in (Rogova et al., 2006).

and relationships characterizing the situation and related to the assumptions, or defined by linguistic labels (*low, medium, high*) with quantization of these values.

Formally, let $\Theta = \{\theta_1^t, \dots, \theta_K^t\}$ be a set of hypotheses under consideration. $Bel(\emptyset) \neq 0$ because, according to the open world assumption, this set of hypothesis is not exhaustive. BAS is a tuple (A,P, ξ,B), in which, as in PAS, A=$\{a_j\}$ is a set of uncertain assumptions, P=$\{p_i\}$ is the set of propositions, and $\xi \in L_{P \cup A}$ is a knowledge base representing a set of rules. At the same time, unlike to PAS, B=$\{bel_j\}$ bel are non-additive dynamic beliefs associated with A=$\{a_j\}$. Arguments Arg_{k_m} supporting (or refuting) each hypothesis θ_k are derived from the knowledge base, and are a conjunction of propositions and assumptions for which θ_k becomes true (or false): $Arg_n(\theta_k) = \bigwedge_j a_{n_j} \bigwedge_k p_{n_k}$. The support of each hypothesis θ_k is defined as the disjunction of all minimal arguments supporting θ_k: Arg(θ_k)=$\bigvee_n ArgP_n$ $\bigvee_m ArgC_m$, where $\bigvee_n ArgP_n$ is a disjunction of all arguments supporting hypothesis θ_k, and $\bigvee_m ArgC_m$ is a disjunction of all arguments refuting hypothesis θ_k.

Beliefs in support of each hypothesis θ_k can be computed by utilizing beliefs in arguments in the following way. Beliefs in support of and against of each assumption a_{n_j} invoke support functions on a frame of discernment Ω_{n_j}=$\{T,F\}$, which have a single focal element (assumption i is true or false). Let us consider a mapping $M : \Omega_{n_1} \times \dots \times \Omega_{n_N} \rightarrow \Theta$. Then, a simple support function μ_k with focus θ_1 in support of argument $ArgP_n$ is:

$$\mu_{ArgP_n}(\theta_k) = \prod_{ArgP_n = \bigwedge_j a_{n_j}} bpa_{a_{n_j}}(T), \mu_{ArgP_n}(\Theta) = 1 - \mu_{ArgP_n}(\theta_k) \tag{C.1}$$

Analogously, the sum of the support functions over the set $\{\Omega_{m_j} \mid \bigwedge_j a_{m_j} = ArgC_m, \forall m\}$ can be directly mapped into a support function ν_j:

$$\nu_{ArgC_n}(\theta_k) = \prod_{ArgC_m = \bigwedge_j a_{m_j}} bpa_{a_{m_j}}(F), \nu_{ArgC_m}(\Theta) = 1 - \nu_{ArgC_m}(\theta_k) \tag{C.2}$$

Accordingly, arguments pro and contra each hypothesis are used to compute hypothesis belief as a combination of μ_k and ν_j for all k and j with the unnormalized Dempster rule. This result is used for decision state estimation.

As it was mentioned before, the process of hypothesis selection requires consideration of decision quality, which has to be evaluated against time required for additional observations/computations. In addition, decision process on any hypothesis under consideration has to take into account that something totally unexpected and not included in the possible causes of the observed situational elements can happen.

The decision rule considered is the following (Yager, 1987):

- If $Bel^t(\emptyset) \geq \max(Bel^t(A)), \forall A \subseteq \Theta$ (i.e., the level of support for an unknown hypothesis exceeds the level of support for any hypothesis under consideration), then the expert operator is alerted to reassess the considered hypotheses set. Additionally, a sensor management process can be started to verify and improve the incoming information.

- Otherwise,
  - If $Bel^t(\Theta) \geq max(Bel^t(A)), \forall A \subseteq \Theta$ (i.e., the level of ignorance exceeds beliefs in any hypothesis), then wait until additional information arrives at the next time step.
  - If $BetP^t(\theta_k) \geq th(t)BetP^t(\theta_n) \forall n \neq k$ then select θ_k, otherwise wait.

$BetP^t(\theta_k)$ is the pignistic probability[8] of hypothesis θ_k at time t=10; *th(t)* is a threshold varying in time that can be modeled by a context-specific decreasing convex function that is set to zero after a certain value.

[8] The term pignistic was coined by C.A.B. Smith (Smith, 1961) from pignus –a bet in Latin– to define a probability function constructed from a belief function for decision–making.

References

Albusac, J., Vallejo, D., Castro-Schez, J. J., Remagnino, P., Gonzalez, C., and Jimenez, L. (2010). Monitoring complex environments using a knowledge-driven approach based on intelligent agents. *IEEE Intelligent Systems*, (3):24–31.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843.

Augusto, J. C. (2007). Ambient intelligence: the confluence of ubiquitous/pervasive computing and artificial intelligence. In *Intelligent Computing Everywhere*, pages 213–234. Springer.

Baader, F., Knechtel, M., and Peñaloza, R. (2012). Context-dependent views to axioms and consequences of semantic web ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web*, 12:22–40.

Baldauf, M., Dustdar, S., and Rosenberg, F. (2007). A survey on context-aware systems. *International Journal of Ad Hoc and Ubiquitous Computing*, 2(4):263–277.

Bao, J., Tao, J., McGuinness, D. L., and Smart, P. (2010). Context representation for the semantic web. In *Proceedings of the Web Science Conference (WebSci10)*.

Barnett, J. A. (2008). Computational methods for a mathematical theory of evidence. In *Classic Works of the Dempster-Shafer Theory of Belief Functions*, pages 197–216. Springer.

Baumgartner, N., Gottesheim, W., Mitsch, S., Retschitzegger, W., and Schwinger, W. (2010). Beaware!–situation awareness, the ontology-driven way. *Data & Knowledge Engineering*, 69(11):1181–1193.

Bellenger, A., Lerouvreur, X., Abdulrab, H., and Kotowicz, J.-P. (2012). Semantic beliefs fusion. In *Advances on Computational Intelligence*, pages 491–500. Springer.

Benavoli, A., Chisci, L., Farina, A., Timmoneri, L., and Zappa, G. (2006). Knowledge-based system for multi-target tracking in a littoral environment. *IEEE Transactions on Aerospace and Electronic Systems*, 42(3):1100–1119.

Bennett, B. (1994). Spatial reasoning with propositional logics. In *Principles of knowledge representation and reasoning: proceedings of the 4th International Conference (KR'94)*.

Besada, J. A., García, J., Portillo, J., Molina, J. M., Varona, A., Gonzalez, G., et al. (2005). Airport surface surveillance based on video images. *IEEE transactions on aerospace and electronic systems*, 41(3):1075–1082.

Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive and Mobile Computing*, 6(2):161–180.

Bevilacqua, A., Stefano, L. D., and Azzari, P. (2006). People tracking using a Time-of-Flight depth sensor. In *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06)*, pages 89–89, Sydney, Australia.

Bille, W., Troyer, O. D., Kleinermann, F., Pellens, B., and Romero, R. (2004). Using ontologies to build virtual worlds for the web. In *Proceedings of the 3rd IADIS International Conference WWW/Internet (ICWI'04)*, pages 683–690, Madrid, Spain.

Bouquet, P., Giunchiglia, F., van Harmelen, F., Serafini, L., and Stuckenschmidt, H. (2004). Contextualizing ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web*, 1(4):325–343.

Bradley, N. A. and Dunlop, M. D. (2005). Toward a multidisciplinary model of context to support context-aware computing. *Human-Computer Interaction*, 20(4):403–446.

Bremond, F. and Thonnat, M. (1996). A context representation for surveillance systems. *ECCV Worshop on Conceptual Descriptions from Images*, 2.

Breuer, P., Eckes, C., and Müller, S. (2007). Hand gesture recognition with a novel IR Time-of-Flight range camera–a pilot study. In *Proceedings of the 3rd International Conference on Computer Vision/Computer Graphics Collaboration Techniques (Mirage'07)*, volume 4418, pages 247–260, Rocquencourt, France.

Carvalho, R. N., Haberlin, R., Costa, P. C. G., Laskey, K. B., and Chang, K.-C. (2011). Modeling a probabilistic ontology for maritime domain awareness. In *Proceedings of the 14th International Conference on Information Fusion (FUSION)*, pages 1–8.

Carvalho, R. N., Laskey, K. B., and Costa, P. C. (2013). Pr-owl 2.0–bridging the gap to owl semantics. In *Uncertainty Reasoning for the Semantic Web II*, pages 1–18. Springer.

Chen, H., Finin, T., and Joshi, A. (2004a). Semantic web in the context broker architecture. In *Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04)*, page 277.

Chen, H., Perich, F., Finin, T., and Joshi, A. (2004b). SOUPA: Standard ontology for ubiquitous and pervasive applications. In *The First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004.*, pages 258–267.

Chen, W., Kifer, M., and Warren, D. S. (1993). HiLog: A foundation for higher-order logic programming. *The Journal of Logic Programming*, 15(3):187–230.

Clark, B. (1985). Individuals and points. *Notre Dame Journal of Formal Logic*, 26(1):61–75.

Clarke, B. (1981). A calculus of individuals based on "connection". *Notre Dame Journal of Formal Logic*, 22(3):204–218.

Cohn, A. G., Magee, D. R., Galata, A., Hogg, D. C., and Hazarika, S. M. (2003). Towards an architecture for cognitive vision using qualitative spatio-temporal representations and abduction. *Spatial cognition III*, 3:232–248.

Corchado, J. M., Bajo, J., and Abraham, A. (2008). Gerami: Improving healthcare delivery in geriatric residences. *IEEE Intelligent Systems*, 23(2):19–25.

Costa, P. C. (2005). *Bayesian semantics for the Semantic Web.* George Mason University.

Costa, P. C., Ladeira, M., Carvalho, R. N., Laskey, K. B., Santos, L. L., and Matsumoto, S. (2008). A first-order bayesian tool for probabilistic ontologies. *Proceedings of the 21st Florida FLAIRS UNBBayes-MEBN.*

Costa, P. C., Laskey, K. B., Blasch, E., and Jousselme, A.-L. (2012). Towards unbiased evaluation of uncertainty reasoning: the urref ontology. In *15th International Conference on Information Fusion*, pages 2301–2308. IEEE.

Coutaz, J., Crowley, J., Dobson, S., and Garlan, D. (2005). Context is key. *Communications of the ACM*, 48(3):49–53.

Dey, A. (1998). Context-aware computing: The cyberdesk project. In *Proceedings of the AAAI, Spring Symposium on Intelligent Environments*, Menlo Park, California, USA.

Dey, A. and Abowd, G. (2000). Towards a better understanding of context and context-awareness. In *Proceedings of the Workshop on the what, who, where, when, why and how of context-awareness*.

Ding, Z. and Peng, Y. (2004). A probabilistic extension to ontology language owl. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04)*.

Doermann, D. and Mihalcik, D. (2000). Tools and techniques for video performance evaluation. In *15th International Conference on Pattern Recognition*. IEEE.

Egenhofer, M. (1991). Reasoning about binary topological relations. In *Proceedings of the Second International Symposium on Advances in Spatial Databases*, pages 143–160.

Elsenbroich, C., Kutz, O., and Sattler, U. (2006). A case for abductive reasoning over ontologies. In *OWL workshop: experiences and directions. (OWLED)*, volume 216.

Essaid, A., Ben Yaghlane, B., and Mellouli, K. (2012). BeliefOWL: An evidential representation in OWL ontology. *Communications of the ACS*, 4(2).

Fagin, R., Halpern, J. Y., and Megiddo, N. (1990). A logic for reasoning about probabilities. *Information and computation*, 87(1):78–128.

Fischer, G. (2012). Context-aware systems: the 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 287–294. ACM.

Foix, S., Alenyà, G., and Torras, C. (2010). Exploitation of Time-of-Flight (ToF) cameras. IRI Tech. Rep. 10-07, Institut de Robótica i Informática Industrial, CSIC-UPC, Barcelona, Spain.

Foix, S., Alenya, G., and Torras, C. (2011). Lock-in Time-of-Flight (ToF) cameras: A survey. *IEEE Sensors Journal*, 11:1917–1926.

Ganapathi, V., Plagemann, C., Koller, D., and Thrun, S. (2010). Real time motion capture using a single time-of-flight camera. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 755–762.

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). Sweetening ontologies with DOLCE. In *Knowledge Engineering and Knowledge Management. 13th International Conference on Ontologies and the Semantic Web (EKAW'02)*, pages 223–233, Sigüenza, Spain.

García, J., Gómez-Romero, J., Patricio, M. A., Molina, J. M., and Rogova, G. (2011). On the representation and exploitation of context knowledge in a harbor surveillance scenario. In *14th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE.

García, J., Molina, J. M., Besada, J. A., and Portillo, J. I. (2005). A multitarget tracking video system based on fuzzy and neuro-fuzzy techniques. *EURASIP Journal on Applied Signal Processing*, 14:2341–2358.

García, J., Molina, J. M., and Llinas, J. (2012). High-level information fusion in visual sensor networks. *Visual Information Processing in Wireless Sensor Networks*.

George, J., Crassidis, J. L., and Singh, T. (2009). Threat assessment using context-based tracking in a maritime environment. In *12th International Conference on Information Fusion*, pages 187–194. IEEE.

Girju, R., Badulescu, A., and Moldovan, D. (2006). Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135.

Giunchiglia, F. (1993). Contextual reasoning. *Epistemologia, special issue on I Linguaggi e le Macchine*, 16:345–364.

Gokturk, S. B. and Tomasi, C. (2004). 3d head tracking based on recognition and interpolation using a Time-of-Flight depth sensor. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, volume 2, pages 211–217, Washington, USA.

Gómez-Romero, J., García, J., Kandefer, M., Llinas, J., Molina, J. M., Patricio, M. A., Prentice, M., and Shapiro, S. C. (2010). Strategies and techniques for use and exploitation of contextual information in high-level fusion architectures. In *13th Conference on Information Fusion*, pages 1–8. IEEE.

Gómez-Romero, J., Patricio, M., García, J., and Molina, J. M. (2009). Towards the implementation of an ontology-based reasoning system for visual information fusion. In *3rd Annual Skövde Workshop on Information Fusion Topics (SWIFT 2009)*, pages 5–10.

Gómez-Romero, J., Patricio, M. Á., García, J., and Molina, J. M. (2011a). Communication in distributed tracking systems: an ontology-based approach to improve cooperation. *Expert Systems*, 28(4):288–305.

Gómez-Romero, J., Patricio, M. A., García, J., and Molina, J. M. (2011b). Ontology-based context representation and reasoning for object tracking and scene interpretation in video. *Expert Systems with Applications*, 38(6):7494–7510.

Gómez-Romero, J., Serrano, M. A., Patricio, M. A., García, J., and Molina, J. M. (2011c). Context-based scene recognition from visual data in smart homes: An Information Fusion approach. *Personal and Ubiquitous Computing*, pages 1–23.

Gómez Romero, J., Serrano, M. Á., Patricio Guisado, M. Á., and Molina, J. M. (2013). A practical approach to the development of ontology-based information fusion systems. *Prediction and recognition of piracy efforts using collaborative human-centric information systems*.

Gong, L. (2005). Contextual modeling and applications. In *IEEE International Conference on Systems, Man and Cybernetics*, volume 1, pages 381–386. IEEE.

Gordon, J. and Shortliffe, E. H. (1985). A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26(3):323–357.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220.

Grutter, R. and Bauer-Messmer, B. (2007a). Combining OWL with RCC for spatioterminological reasoning on environmental data. In *Proceedings of 3rd OWL: Experiences and Directions (OWLED'07)*.

Grutter, R. and Bauer-Messmer, B. (2007b). Towards spatial reasoning in the semantic web: A hybrid knowledge representation system architecture. *The European Information Society*, pages 349–364.

Grütter, R., Scharrenbach, T., and Bauer-Messmer, B. (2008). Improving an rcc-derived geospatial approximation by owl axioms. In *7th International Semantic Web Conference (ISWC 2008)*, pages 293–306. Springer.

Gu, T., Pung, H. K., Zhang, D. Q., Pung, H. K., and Zhang, D. Q. (2004a). A bayesian approach for dealing with uncertain contexts. In *Austrian Computer Society*.

Gu, T., Wang, X. H., Pung, H. K., and Zhang, D. Q. (2004b). An ontology-based context model in intelligent environments. In *Proceedings of communication networks and distributed systems modeling and simulation conference*, volume 2004, pages 270–275.

Guard, U. C. (2007). The us coast guard strategy for maritime safety, security, and stewardship. *Retrieved March*.

Guerriero, M., Willett, P., Coraluppi, S., and Carthel, C. (2008). Radar/ais data fusion and sar tasking for maritime surveillance. In *11th International Conference on Information Fusion*, pages 1–5. IEEE.

Guha, R., McCool, R., and Fikes, R. (2004). Contexts for the semantic web. In *The Semantic Web (ISWC'04)*, pages 32–46. Springer.

Guizzardi, G. (2005). *Ontological foundations for structural conceptual models*. CTIT, Centre for Telematics and Information Technology.

Guomundsson, S. A., Larsen, R., Aanaes, H., Pardas, M., and Casas, J. R. (2008). ToF imaging in smart room environments towards improved people tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)*, pages 1–6, Anchorage, USA.

Gustavsen, R. (2002). Condor–an application framework for mobility-based context-aware applications. In *Proceedings of the workshop on concepts and models for ubiquitous computing*, volume 39.

Gutiérrez, M., Garcia-Rojas, A., Thalmann, D., Vexo, F., Moccozet, L., Magnenat-Thalmann, N., Mortara, M., and Spagnuolo, M. (2007). An ontology of virtual humans. *The Visual Computer*, 23(3):207–218.

Haarslev, V. and Möller, R. (2001). Description of the racer system and its applications. In *International workshop on description logics (DL2001)*, volume 49.

Haenni, R., Kohlas, J., and Lehmann, N. (2000). *Probabilistic argumentation systems*, volume 5. Springer.

Haker, M., Bohme, M., Martinetz, T., and Barth, E. (2007). Geometric invariants for facial feature tracking with 3D ToF cameras. In *International Symposium on Signals, Circuits and Systems (ISSCS'07)*, volume 1, pages 1–4, Iasi, Romania.

Hall, D. and Llinas, J. (2009). *Multisensor data fusion*. CRC press.

Hansen, D. W., Larsen, R., and Lauze, F. (2007). Improving face detection with ToF cameras. In *Proceedings of the International Symposium on Signals, Circuits and Systems (ISSCS'07)*, volume 1, pages 1–4, Iasi, Romania.

Harter, A. and Hopper, A. (1994). A distributed location system for the active office. *Network, IEEE*, 8(1):62–70.

Heinsohn, J. (1994). Probabilistic description logics. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence*, pages 311–318.

Henderson-Sellers, B. and Barbier, F. (1999). What is this thing called aggregation? In *Proceedings of Technology of Object-Oriented Languages and Systems*, pages 236–250.

Henricksen, K. (2003). *A framework for context-aware pervasive computing applications. Ph.D. thesis*. University of Queensland Queensland.

Henricksen, K. and Indulska, J. (2004). A software engineering framework for context-aware pervasive computing. In *Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications. (PerCom'04)*, pages 77–86, Orlando, USA.

Henricksen, K. and Indulska, J. (2006). Developing context-aware pervasive computing applications: Models and approach. *Pervasive and mobile computing*, 2(1):37–64.

Henricksen, K., Livingstone, S., and Indulska, J. (2004). Towards a hybrid approach to context modeling, reasoning and interoperation. In *The First International Workshop on Advanced Context Modelling, Reasoning and Management*, pages 54–61.

Hofer, T., Schwinger, W., Pichler, M., Leonhartsberger, G., Altmann, J., and Retschitzegger, W. (2002). Context-awareness on mobile devices–the hydrogen approach. In *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, pages 292–302.

Hogenboom, F., Borgman, B., Frasincar, F., and Kaymak, U. (2010). Spatial knowledge representation on the semantic web. In *IEEE Fourth International Conference on Semantic Computing (ICSC)*, pages 252–259.

Hois, J. (2007). Towards combining ontologies and uncertain knowledge. In *Proc. 3rd Workshop on Combining Probability and Logic*.

Hölldobler, S., Khang, T. D., and Störr, H.-P. (2002). A fuzzy description logic with hedges as concept modifiers. In *Proceedings InTech/VJFuzzy*, volume 2002, pages 25–34.

Holte, M. B., Moeslund, T. B., and Fihl, P. (2008). Fusion of range and intensity information for view invariant gesture recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)*, pages 1–7, Anchorage, USA.

Hong, J.-y., Suh, E.-h., and Kim, S.-J. (2009). Context-aware systems: A literature review and classification. *Expert Systems with Applications*, 36(4):8509–8522.

Hornsby, K. S. and Joshi, K. (2010). Combining ontologies to automatically generate temporal perspectives of geospatial domains. *Geoinformatica*, 14(4):481–505.

Horridge, M. and Bechhofer, S. (2009). The owl api: A java api for working with owl 2 ontologies. In *6th OWL Experienced and Directions Workshop (OWLED 2009)*.

Horrocks, I. (2002). DAML+OIL: A description logic for the semantic web. *Bulletin of the Technical Committee on Data Engineering*, 25(1):4–9.

Horrocks, I., Kutz, O., and Sattler, U. (2006). The even more irresistible SROIQ. In *Proceedings of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR'06)*, pages 57–67.

Horrocks, I. and Patel-Schneider, P. F. (2004). A proposal for an owl rules language. In *13th international conference on World Wide Web (WWW 2004)*, pages 723–731. ACM.

Isard, M. and Blake, A. (1998). Condensation–conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28.

Josephson, J. R. (1990). On the logical form of abduction. In *AAAI Spring Symposium Series: Automated Abduction*, pages 140–144.

Kahlmann, T., Remondino, F., and Guillaume, S. (2007). Range imaging technology: New developments and applications for people identification and tracking. In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE'07)*, volume 6491.

Kandefer, M. and Shapiro, S. C. (2008). A categorization of contextual constraints. In *AAAI Fall Symposium: Biologically Inspired Cognitive Architectures*, pages 88–93.

Kappel, G., Retschitzegger, W., Kimmerstorfer, E., B., P., Schwinger, W., and Hofer, T. (2002). Towards a generic customisation model for ubiquitous web applications. In *2nd International Workshop on Web Oriented Software Technology (IWWOST'2002)*.

Katz, Y. and Grau, B. (2005a). BC: Representing qualitative spatial information in OWL-DL. In *Proceedings of OWL: Experiences and Directions*.

Katz, Y. and Grau, B. C. (2005b). Representing qualitative spatial information in owl-dl. In *OWL: Experiences and Directions Workshop (OWLED 2005)*.

Keet, C. M. and Artale, A. (2008). Representing and reasoning over a taxonomy of part-whole relations. *Applied Ontology*, 3(1):91–110.

Klien, E., Lutz, M., and Kuhn, W. (2006). Ontology-based discovery of geographic information services–An application in disaster management. *Computers, environment and urban systems*, 30(1):102–123.

Knoop, S., Vacek, S., and Dillmann, R. (2005). Modeling joint constraints for an articulated 3D human body model with artificial correspondences in ICP. In *5th IEEE-RAS International Conference on Humanoid Robots*, pages 74–79, Tsukuba, Japan.

Knoop, S., Vacek, S., and Dillmann, R. (2006). Sensor fusion for 3D human body tracking with an articulated 3D body model. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'06)*, pages 1686–1691, Orlando, USA.

Kokar, M. M., Matheus, C. J., and Baclawski, K. (2009). Ontology-based situation awareness. *Information fusion*, 10(1):83–98.

Kolb, A., Barth, E., Koch, R., and Larsen, R. (2009). Time-of-Flight sensors in computer graphics. *Eurographics State of the Art Reports*, pages 119–134.

Koller, D., Levy, A., and Pfeffer, A. (1997). P-classic: A tractable probablistic description logic. In *Proceedings of the National Conference on Artificial Intelligence*, pages 390–397.

Laskey, K. B., Haberlin, R., Carvalho, R. N., and da Costa, P. C. G. (2011). Pr-owl 2 case study: A maritime domain probabilistic ontology. In *6th International Conference on Semantic Technologies for Intelligence, Defense and Security (STIDS 2011)*, pages 76–83.

Leone, A., Diraco, G., and Siciliano, P. (2011). Detecting falls with 3D range camera in ambient assisted living applications: A preliminary study. *Medical Engineering & Physics*, 33(6):770–781.

Li, Y., Xu, B., Lu, J., and Kang, D. (2006). Discrete tableau algorithms for SHI. In *Proceeedings of the International Workshop on Description Logics (DL-06)*.

Liebhaber, M. J. and Feher, B. (2002). Surface warfare threat assessment: Requirements definition. Technical report, DTIC Document.

Little, E. G. and Rogova, G. L. (2009). Designing ontologies for higher level fusion. *Information Fusion*, 10(1):70–82.

Llinas, J., Bowman, C., Rogova, G., Steinberg, A., et al. (2004). Revisiting the jdl data fusion model ii. In *Proceedings of the 7th International Conference on Information Fusion (FUSION 2004)*.

Lukasiewicz, T. and Straccia, U. (2008). Managing uncertainty and vagueness in description logics for the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(4):291–308.

Maillot, N., Thonnat, M., and Boucher, A. (2004). Towards ontology-based cognitive vision. *Machine Vision and Applications*, 16(1):33–40.

Matheus, C. J., Kokar, M. M., Baclawski, K., Letkowski, J. A., Call, C., Hinman, M. L., Salerno, J. J., and Boulware, D. M. (2005). Sawa: An assistant for higher-level fusion and situation awareness. In *Defense and Security*, pages 75–85.

McCarthy, J. (1993). Notes on formalizing context. In *Proceedings of the 13th International Conference on Artificial Intelligence (IJCAI'93)*, pages 555–562.

McCarthy, J. and Buvac, S. (1997). Formalizing context (expanded notes). In *Working Papers of the AAAI Fall Symposium on Context in Knowledge Representation and Natural Language*, pages 99–135.

Motik, B. (2012). Representing and querying validity time in rdf and owl: A logic-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 12:3–21.

Motik, B., Sattler, U., and Studer, R. (2005). Query answering for owl-dl with rules. *Web Semantics: Science, Services and Agents on the World Wide Web*, 3(1):41–60.

Neumann, B. and Möller, R. (2008). On scene interpretation with description logics. *Image and Vision Computing*, 26(1):82–101.

Nicholson, A. E. and Brady, J. M. (1994). Dynamic belief networks for discrete monitoring. *IEEE Transactions on Systems, Man and Cybernetics*, 24(11):1593–1610.

Nowak, C. (2003). On ontologies for high-level information fusion. In *Proceedings of the Sixth International Conference of Information Fusion*, volume 1, pages 657–664.

Odell, J. J. (1994). Six different kinds of composition. *Journal of Object-Oriented Programming*, 5(8):10–15.

Opdahl, A. L., Henderson-Sellers, B., and Barbier, F. (2001). Ontological analysis of whole–part relationships in OO-models. *Information and Software Technology*, 43(6):387–399.

Pan, J. Z., Stamou, G., Stoilos, G., Taylor, S., and Thomas, E. (2008). Scalable querying services over fuzzy ontologies. In *Proceedings of the 17th international conference on World Wide Web*, pages 575–584.

Pascoe, J. (1998). Adding generic contextual capabilities to wearable computers. In *Second International Symposium on Wearable Computers*, pages 92–99.

Patricio, M. A., Carbó, J., Pérez, O., García, J., and Molina, J. M. (2007). Multi-agent framework in visual sensor networks. *EURASIP Journal on Applied Signal Processing*, 2007(1):226–247.

Pearl, J. (1988a). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausble Inference*. Morgan Kaufmann Pub.

Pearl, J. (1988b). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

Pinz, A., Bischof, H., Kropatsch, W., Schweighofer, G., Haxhimusa, Y., Opelt, A., and Ion, A. (2008). Representations for cognitive vision: A review of appearance-based, spatio-temporal, and graph-based approaches. *ELCVIA*, 7(2).

Pinz, A., Bischof, H., Kropatsch, W., Schweighofer, G., Haxhimusa, Y., Opelt, A., and Ion, A. (2009). Representations for cognitive vision: A review of appearance-based, spatio-temporal, and graph-based approaches. *Electronic Letters on Computer Vision and Image Analysis (ELCVIA)*, 7(2):35–61.

Pinza, A., Ionc, A., Kropatschc, W., Haxhimusad, Y., Schweighofera, G., Opelta, A., and Bischofb, H. (2008). Representations for cognitive vision: A review of appearance-based, spatio-temporal, and graph-based approaches. In *ELCVIA: Electronic Letters on Computer Vision and Image Analysis*, volume 7, pages 35–61.

Pool, M. and Aikin, J. (1994). Keeper and protégé: An elicitation environment for bayesian inference tools. In *Proceedings of the Workshop on Protégé and Reasoning, held at the 7th International Protégé Conference*.

Powell, G. M., Kokar, M. M., Matheus, C. J., and Lorenz, D. (2006). Understanding the role of context in the interpretation of complex battlespace intelligence. In *9th International Conference on Information Fusion*, pages 1–8. IEEE.

Randell, D., Cui, Z., and Cohn, A. (1992). A spatial logic based on regions and connection. In *4th International Conference on Principles of knowledge representation and reasoning (KR'94)*, volume 92, pages 165–176.

Ranganathan, A., Al-Muhtadi, J., and Campbell, R. H. (2004). Reasoning about uncertain contexts in pervasive computing environments. *Pervasive Computing, IEEE*, 3(2):62–70.

Ranganathan, A., McGrath, R. E., Campbell, R. H., and Mickunas, M. D. (2003). Use of ontologies in a pervasive computing environment. *The Knowledge Engineering Review*, 18(3):209–220.

Rector, A., Welty, C., Noy, N., and Wallace, E. (2005). Simple part-whole relations in OWL ontologies. W3C working draft – work in progress, W3C.

Remagnino, P. and Foresti, G. L. (2005). Ambient intelligence: A new multidisciplinary paradigm. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 35(1):1–6.

Remagnino, P. and Foresti, G. L. (2009). Computer vision methods for ambient intelligence. *Image and Vision Computing*, 27(10):1419–1420.

Renz, J. (2002). *Qualitative spatial reasoning with topological information*. Springer-Verlag.

Riboni, D. and Bettini, C. (2011). OWL 2 modeling and reasoning with complex human activities. *Pervasive and Mobile Computing*, 7(3):379–395.

Rogova, G. L., Scott, P. D., Lollett, C., and Mudiyanur, R. (2006). Reasoning about situations in the early post-disaster response environment. In *9th International Conference on Information Fusion (FUSION 2006)*, pages 1–8. IEEE.

Rosse, C. and Jr, J. L. V. M. (2003). A reference ontology for biomedical informatics: The foundational model of anatomy. *Journal of Biomedical Informatics*, 36(6):478–500.

Roy, J. and Davenport, M. (2010). Exploitation of maritime domain ontologies for anomaly detection and threat analysis. In *2nd International on Waterside Security Conference (WSS 2010)*, pages 1–8. IEEE.

Rudolph, S., Krötzsch, M., and Hitzler, P. (2008). Cheap boolean role constructors for description logics. *Logics in Artificial Intelligence*, pages 362–374.

Ryan, N., Pascoe, J., and Morse, D. (1997). Enhanced reality fieldwork: the context-aware archaeological assistant. In *Computer applications in archaeology*.

Sánchez, D. and Tettamanzi, A. (2005). Generalizing quantification in fuzzy description logics. *Computational Intelligence, Theory and Applications*, pages 397–411.

Sánchez, D. and Tettamanzi, A. G. (2006). Fuzzy quantification in fuzzy description logics. *Capturing Intelligence*, 1:135–159.

Sattler, U. (1995). A concept language for an engineering application with part-whole relations. In *Proceedings of the International Workshop on Description Logics*, pages 119–123.

Schilit, B., Adams, N., and Want, R. (1994). Context-aware computing applications. In *First Workshop on Mobile Computing Systems and Applications, WMCSA.*, pages 85–90.

Schilit, B. and Theimer, M. (1994). Disseminating active map information to mobile hosts. *Network, IEEE*, 8(5):22–32.

Schilit, B., Theimer, M., and Welch, B. (1993). Customizing mobile applications. In *Proceedings USENIX Symposium on Mobile & Location-indendent Computing*, pages 129–138.

Schill, K., Zetzsche, C., and Hois, J. (2009). A belief-based architecture for scene analysis: From sensorimotor features to knowledge and ontology. *Fuzzy Sets and Systems*, 160(10):1507–1516.

Schuele, M. and Karaenke, P. (2010). Qualitative spatial reasoning with topological information in bdi agents. In *Proceedings of the 2nd Workshop on Artificial Intelligence and Logistics (AILog)*, pages 7–12.

Seibert, M., Rhodes, B. J., Bomberger, N. A., Beane, P. O., Sroka, J. J., Kogel, W., Kreamer, W., Stauffer, C., Kirschner, L., Chalom, E., et al. (2006). Seecoast port surveillance. In *SPIE 6204: Photonics for Port and Harbor Security*.

Serafini, L. and Bouquet, P. (2004). Comparing formal theories of context in ai. *Artificial intelligence*, 155(1):41–67.

Serrano, M. A., Gómez-Romero, J., Patricio, M. A., Garcia, J., and Molina, J. M. (2011). Topological properties in ontology-based applications. In *11th International Conference on Intelligent Systems Design and Applications (ISDA)*, pages 1329–1334.

Serrano, M. A., Gracía, J., Patricio, M. A., and Molina, J. M. (2010). Interactive video annotation tool. In *International Symposium on Distributed Computing and Artificial Intelligence (DCAI'10)*, pages 325–332. Springer.

Shafer, G. (1976). *A mathematical theory of evidence*, volume 1. Princeton university press.

Sheng, Q. Z. and Benatallah, B. (2005). Contextuml: a uml-based modeling language for model-driven development of context-aware web services. In *International Conference on Mobile Business, 2005. ICMB 2005.*, pages 206–212.

Simons, P. (1987). *Parts: A study in ontology*. Oxford university press.

Simons, P. (1991). Part/whole II: Mereology since 1900. *Handbook of Metaphysics and Ontology.*, pages 672–675.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y. (2007). Pellet: A practical owl-dl reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):51–53.

Smets, P. (1990). The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458.

Smets, P. and Kennes, R. (1994). The transferable belief model. *Artificial intelligence*, 66(2):191–234.

Smith, C. A. (1961). Consistency in statistical inference and decision. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–37.

Snidaro, L. and Foresti, G. L. (2007). Knowledge representation for ambient security. *Expert Systems*, 24(5):321–333.

Snidaro, L., Visentini, I., and Bryan, K. (2015). Fusing uncertain knowledge and evidence for maritime situational awareness via markov logic networks. *Information Fusion*, 21:159–172.

Snidaro, L., Visentini, I., Bryan, K., and Foresti, G. L. (2012). Markov logic networks for context integration and situation assessment in maritime domain. In *15th International Conference on Information Fusion (FUSION)*, pages 1534–1539.

Soutschek, S., Penne, J., Hornegger, J., and Kornhuber, J. (2008). 3-D gesture-based scene navigation in medical imaging applications using Time-of-Flight cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)*, volume 1, pages 1–6, Anchorage, USA.

Sowa, J. F. (1995). Syntax, semantics, and pragmatics of contexts. In *Proceedings of the 3rd International Conference on Conceptual Structures (ICCS'95)*, pages 1–22. Springer.

Spreitzer, M. and Theimer, M. (1993). Providing location information in a ubiquitous computing environment. In *Proceedings of the Fourteenth ACM Symposium on Operating System Principles*, pages 270–283.

Springer, T. and Turhan, A.-Y. (2009). Employing description logics in ambient intelligence for modeling and reasoning about complex situations. *Journal of Ambient Intelligence and Smart Environments (JAISE)*, 1(3):235–259.

Steinberg, A. N. and Bowman, C. L. (2004). Rethinking the jdl data fusion levels. *MSS national symposium on sensor and data fusion*.

Steinberg, A. N., Bowman, C. L., and White, F. E. (1999). Revisions to the jdl data fusion model. In *Sensor Fusion: Architectures, Algorithms, and Applications III*, volume 3719, pages 430–441.

Steinberg, A. N. and Rogova, G. (2008). Situation and context in data fusion and natural language understanding. In *11th International Conference on Information Fusion*, pages 1–8. IEEE.

Stocker, M. and Sirin, E. (2009). Pelletspatial: A hybrid RCC-8 and RDF/OWL reasoning and query engine. In *CEUR Workshop Proceedings*, volume 529, pages 2–31.

Stoilos, G., Stamou, G., Pan, J. Z., Tzouvaras, V., and Horrocks, I. (2007). Reasoning with very expressive fuzzy description logics. *Journal of Artificial Intelligence Research*, 30(5):273–320.

Stoilos, G., Stamou, G., Tzouvaras, V., Pan, J. Z., and Horrocks, I. (2005). Fuzzy owl: Uncertainty and the semantic web. In *Proceedings of the International Workshop on OWL: Experiences and Directions*, volume 280.

Straccia, U. (1998). A fuzzy description logic. In *Proceedings of the national conference on Artificial Intelligence*, pages 594–599.

Straccia, U. (2001). Reasoning within fuzzy description logics. *Journal of Artificial Intelligence Research*, 14:137–166.

Straccia, U. (2005). Towards a fuzzy description logic for the semantic web. In *Proceedings of the Second European conference on The Semantic Web: Research and Applications*, pages 167–181.

Strang, T. and Linnhoff-Popien, C. (2004). A context modeling survey. In *1st International Workshop on Advanced Context Modeling, Reasoning and Management*.

Sycara, K., Glinton, R., Yu, B., Giampapa, J., Owens, S., Lewis, M., and Grindle, L. C. (2009). An integrated approach to high-level information fusion. *Information Fusion*, 10(1):25–50.

Thagard, P. and Shelley, C. (1997). Abductive reasoning: Logic, visual thinking, and coherence. In *Logic and scientific methods*, pages 413–427. Springer.

Tresp, C. B. and Molitor, R. (1998). A description logic for vague knowledge. In *Proceedings of the 13th European Confeference on Artificial Intelligence (ECAI-98*.

Turaga, P., Ivanov, Y., et al. (2011). Diamond sentry: integrating sensors and cameras for real-time monitoring of indoor spaces. *IEEE Sensors Journal*, 11(3):593–602.

Vallejo, D., Albusac, J., Jimenez, L., Gonzalez, C., and Moreno, J. (2009). A cognitive surveillance system for detecting incorrect traffic behaviors. *Expert Systems with Applications*, 36(7):10503–10511.

Van den Broek, A., Neef, R., Hanckmann, P., van Gosliga, S. P., and Van Halsema, D. (2011). Improving maritime situational awareness by fusing sensor information and intelligence. In *Proceedings of the 14th International Conference on Information Fusion*, pages 1–8. IEEE.

Van Hage, W., Wielemaker, J., and Schreiber, G. (2010). The space package: Tight integration between space and semantics. *Transactions in GIS*, 14(2):131–146.

Varzi, A. (2011). Mereology. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Spring 2011 edition.

Varzi, A. C. (1996). Parts, wholes, and part-whole relations: The prospects of mereotopology. *Data & Knowledge Engineering*, 20(3):259–286.

Velastin, S., Boghossian, B., Lo, B. P. L., Sun, J., Vicencio-Silva, M. A., et al. (2005). Prismatica: toward ambient intelligence in public transport environments. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 35(1):164–182.

Vernon, D. (2008). Cognitive vision: The case for embodied perception. *Image and Vision Computing*, 26(1):127–140.

Villaroman, N., Rowe, D., and Swan, B. (2011). Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor. In *Proceedings of the 12th Conference on Information Technology Education (SIGITE'11)*, pages 227–232, New York, USA.

Vinciarelli, A., Pantic, M., and Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, 27(12):1743–1759.

Want, R., Hopper, A., Falcão, V., and Gibbons, J. (1992). The active badge location system. *ACM Transactions on Information Systems*, 10(1):91–102.

Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3):94–104.

Whitehead, A. (1929). *Process and reality*. Simon and Schuster.

Wiegand, N. and García, C. (2007). A task-based ontology approach to automate geospatial data retrieval. *Transactions in GIS*, 11(3):355–376.

Wientapper, F., Ahrens, K., Wuest, H., and Bockholt, U. (2009). Linear-projection-based classification of human postures in Time-of-Flight data. In *IEEE International Conference on Systems, Man and Cybernetics (SMC'09)*, pages 559–564, San Antonio, USA.

Winston, M. E., Chaffin, R., and Herrmann, D. (1987). A taxonomy of part-whole relations. *Cognitive science*, 11(4):417–444.

Wu, C. and Aghajan, H. (2011). User-centric environment discovery with camera networks in smart homes. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 41(2):375–383.

Yager, R. R. (1987). On the Dempster-Shafer framework and new combination rules. *Information sciences*, 41(2):93–137.

Yaghlane, B. B. and Mellouli, K. (2008). Inference in directed evidential networks based on the transferable belief model. *International Journal of Approximate Reasoning*, 48(2):399–418.

Yaghlane, B. B., Smets, P., and Mellouli, K. (2003). Directed evidential networks with conditional belief functions. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 291–305. Springer.

Yang, M., Wu, Y., and Hua, G. (2009). Context-aware visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(7):1195–1209.

Yelland, P. M. (2000). An alternative combination of bayesian networks and description logics. In *Principles of Knowledge Representation and reasoning international conference*, pages 225–234.

Yen, J. (1991). Generalizing term subsumption languages to fuzzy logic. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 472–477.

Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys (CSUR)*, 38(4):13.

Yu, Z., Zhou, X., Yu, Z., Park, J. H., and Ma, J. (2008). iMuseum: A scalable context-aware intelligent museum system. *Computer Communications*, 31(18):4376–4382.

Yujie, Z. and Licai, W. (2010). Some challenges for context-aware recommender systems. In *2010 5th International Conference on Computer Science and Education (ICCSE)*, pages 362–365.

Zadeh, L. A. (1999). Fuzzy sets as a basis for a theory of possibility. *Fuzzy sets and systems*, 100:9–34.

Zhang, D., Cao, J., Zhou, J., and Guo, M. (2009). Extended dempster-shafer theory in context reasoning for ubiquitous computing environments. In *International Conference on Computational Science and Engineering (CSE'09)*, volume 2, pages 205–212.

Zhang, D., Gu, T., and Wang, X. H. (2005). Enabling context-aware smart home with semantic web technologies. *International Journal of Human-friendly Welfare Robotic Systems*, 6(4):12–20.